Métodos para el análisis de los procesos de ciencia, tecnología e innovación

Herramientas para el estudio del desarrollo de América Latina

Volumen 3. **Métodos mixtos y emergentes**

José Miguel Natera
y Diana Suárez
(compiladores)

Colección **Ciencia, innovación y desarrollo**

EDICIONES **UNGS**
Universidad Nacional de General Sarmiento

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

LALICS

Métodos para el análisis de los procesos de ciencia, tecnología e innovación
Herramientas para el estudio del desarrollo de América Latina

Volumen 3. Métodos mixtos y emergentes

Autoras y autores

Diana Suárez: Instituto de Industria - Universidad Nacional de General Sarmiento.

José Miguel Natera: Instituto de Investigaciones Económicas - Universidad Nacional Autónoma de México.

Rodrigo Arocena: Universidad de la República.

Judith Sutz: Universidad de la República.

Nora Goren: Instituto de Estudios Sociales en Contextos de Desigualdades - Universidad Nacional de José C. Paz.

Walter Lugo Ruiz-Castañeda: Universidad Nacional de Colombia.

Juan F. Franco-Bermúdez: Universidad Nacional de Colombia.

María Luisa Villalba-Morales: Universidad Católica de Oriente.

Lilia Stubrin: CONICET - Centro de Investigaciones para la Transformación - Universidad Nacional de San Martin.

Cecilia Tomassini: Comisión Sectorial de Investigación Científica - Universidad de la República.

Ana Urraca Ruiz: Facultad de Ciencias Económicas. Universidade Federal Fluminense.

Pedro Miranda: Instituto de Pesquisa Econômica Aplicada (IPEA).

Vanessa de Lima Avanci: InSySPo - Universidade Estadual de Campinas.

Milton M. Herrera: Centro de Investigaciones en Ciencias Económicas - Universidad Militar Nueva Granada.

Mauricio Uriona Maldonado: Grupo de Sustentabilidad e Innovación en Energías Renovables (Sinergia). Departamento de Ingeniería Industrial y Sistemas - Universidad Federal de Santa Catarina.

Caroline Rodrigues Vaz: Grupo de Sustentabilidad e Innovación en Energías Renovables (Sinergia). Departamento de Ingeniería Industrial y Sistemas - Universidad Federal de Santa Catarina.

William Alejandro Orjuela Garzón: Grupo de investigacion Bioeconomía - Facultad de Ingeniería Agronómica - Univerisdad del Tolima.

Santiago Quintero Ramírez: Escuela de Ingenierías - Universidad Pontificia Bolivariana

Rodrigo Kataishi: CONICET - Universidad Nacional de Tierra del Fuego.

Matías Milia: Posdoctorante, Instituto Tecnológico Autónomo de México.

Maleya Saraí López-Castro: Universidad Complutense de Madrid.

Nayely Martínez: Universidad de Monterrey.

Natalia Gras: Comisión Sectorial de Investigación Científica - Universidad de la República.

José Miguel Natera: Insituto de Investigaciones Económicas - Universidad Nacional Autónoma de México.

José Miguel Natera y Diana Suárez
(compiladores)

Métodos para el análisis de los procesos de ciencia, tecnología e innovación

Herramientas para el estudio del desarrollo de América Latina

Volumen 3. Métodos mixtos y emergentes

EDICIONES UNGS
Universidad Nacional de General Sarmiento

Casa abierta al tiempo
UNIVERSIDAD AUTÓNOMA METROPOLITANA

LALICS

Métodos para el análisis de los procesos de ciencia, tecnología e innovación : herramientas para el estudio del desarrollo de América Latina / José Miguel Natera ... [et al.] ; Editado por José Miguel Natera ; Diana Suárez. - 1a ed - Los Polvorines : Universidad Nacional de General Sarmiento ; Ciudad de México : Universidad Autónoma Metropolitana , 2024.

Libro digital, PDF - (Ciencia, innovación y desarrollo ; 20)

Archivo Digital: descarga y online

ISBN 978-987-630-748-2

1. Innovaciones. 2. América Latina. 3. Desarrollo Tecnológico. I. Natera, José Miguel, ed. II. Suárez, Diana, ed.

CDD 307.14098

EDICIONES **UNGS**

J. M. Gutiérrez 1150, Los Polvorines (B1613GSX)
Prov. de Buenos Aires, Argentina
Tel.: (54 11) 4469-7507
ediciones@ungs.edu.ar
www.ungs.edu.ar/ediciones

Diseño gráfico de la colección: Franco Perticaro
Diseño de tapa: Daniel Vidable
Corrección: Florencia Piluso
Diagramación: Daniel Vidable

Hecho el depósito que marca la Ley 11723.

L U A
Libro
Universitario
Argentino

Índice

Introducción

Diana Suárez y José Miguel Natera

El estudio de los procesos de ciencia, tecnología e innovación (CTI) es un campo consolidado dentro de las ciencias sociales. Su eje está puesto en el análisis de los determinantes, obstáculos, patrones e impacto del desarrollo tecnológico y cómo ello permea (o no) en los procesos de desarrollo. Es un campo con una fuerte raíz teórica en la obra de Schumpeter pero que se ha nutrido de aportes desde múltiples campos del saber, lo que lo constituye en un espacio de debate transdisciplinar tanto en el ámbito académico como de los hacedores de política.

Siendo un espacio de creación de conocimiento científico, el estudio de los procesos de CTI se rige por las normas del método científico, lo que implica que la acumulación de conocimiento tiene lugar a través de la revisión sistemática y crítica del estado del arte, así como de la formulación y contrastación de hipótesis a través de métodos de análisis rigurosos, con la consecuente retroalimentación al estado del arte. Cada una de esas etapas se encuentra más o menos desarrolladas en lo que Cohen y Levinthal (1989) referirían como "el libro de códigos", ese lenguaje compartido, que combina saberes más o menos codificados, y en el marco del cual se construye y deconstruye el conocimiento. Desde hace varias décadas, existen además revistas especializadas, seminarios, conferencias y redes internacionales de colaboración en las que las ideas son sometidas al debate crítico y constructivo, en diálogo además con las múltiples disciplinas que integran el campo. Así, ha habido lugar para que el debate teórico, y su consecuente articulación conceptual, haya tenido un espacio de desarrollo (aún en expansión) permitiendo el abordaje de retos analíticos en América Latina.

En materia de la implementación de métodos para el uso de evidencia empírica, sin embargo, los límites son más difusos. Si bien el corpus conceptual de los estudios CTI se encuentra determinando por la afluencia

de múltiples corrientes del pensamiento, las herramientas metodológicas con las que se pretende observar el objeto de estudio también se encuentran atravesadas por múltiples marcos conceptuales disciplinares, que van desde la economía evolucionista más tradicional hasta nuevos enfoques desde la sociología y la antropología. Así, es posible encontrar trabajos típicamente cuantitativos, con rigurosas técnicas econométricas, análisis cualitativos con técnicas etnográficas y de estudios de caso, y, más recientemente, técnicas mixtas y avanzadas que combinan la capacidad de procesamiento de software especializados con análisis basados en información recogida en entrevistas y observaciones participantes. En este aspecto, el debate ha sido de mucha menor intensidad, reconociendo la validez de las técnicas por su aplicación original (el campo disciplinar de origen) sin demasiada reflexión crítica respecto de los supuestos sobre los que se basa, la posibilidad de extrapolar resultados y la medida en que esos resultados contribuyen a incrementar el cuerpo teórico del campo, en especial cuando se aplica a contextos de menor desarrollo relativo.

A la heterogeneidad y especificidad que supone esta convergencia de campos y métodos, se suma además la complejidad propia del objeto de estudio, a saber, los procesos de CTI. Se trata de un objeto histórico en permanente cambio y con algunos elementos comunes y otros específicos a cada proceso territorial de desarrollo. Más aún, cuando el objeto se acota a un lugar y momento específico, se evidencian particularidades geopolíticas, sociales y culturales, que se suman a procesos también históricos en las estructuras económicas y productivas y, desde luego, de los sistemas nacionales de innovación. Aparece aquí otro desafío para el campo. Requiere de atención a las especificidades de cada territorio para explicar los procesos de CTI, pero también de la generalidad conceptual que permite incrementar nuestro saber respecto del proceso en cuestión.

Frente a ese desafío, los estudios de CTI han probado de manera cabal la necesidad de pensar estos procesos para la región latinoamericana en toda la especificidad intra e interregional, que se replica además en la escala de los países de la región. Estas consideraciones también aplican al plano metodológico, en el que el desafío es además práctico, dado el menor desarrollo de los sistemas de información de la región. Este libro pretende contribuir, en alguna medida, a palear esos desafíos.

El objetivo de esta obra es compilar una serie de métodos y diseños metodológicos empleados para el estudio de los procesos de CTI, con foco en las particularidades y adaptaciones necesarias para su uso en América Latina. Está destinada a personas interesadas en desarrollar estudios en

la temática en la región, particularmente estudiantes de maestría, doctorado e investigadoras/res que deseen explorar nuevas alternativas para el análisis de sus problemas de investigación.

Con este marco, los capítulos organizados en tres volúmenes constituyen un conjunto no exhaustivo de propuestas que han probado ser útiles para el estudio de los procesos de CTI en la región, para un conjunto también no exhaustivo de tópicos sobre los cuales existe un intenso debate académico y de política pública. Asimismo, la selección de estas propuestas ha probado ser útil más allá del tema específico que se propone en cada capítulo, pudiendo servir de referencia para otros problemas de investigación, en especial, a partir de la combinación de capítulos. En este sentido, hemos pretendido que el conjunto de capítulos sea útil, además, para abordar esos temas que hoy no están presentes en el debate pero que por ese objeto de estudio dinámico, histórico y contextual es posible cobren relevancia en un futuro cercano.

Origen y motivación: redes de colaboración y debate regional

Este libro forma parte de las actividades de la Red Latinoamericana para el estudio de los Sistemas de Aprendizaje, Innovación y Construcción de Competencias (Red LALICS). LALICS es una red académica de América Latina y el Caribe que reúne a personas de la comunidad científica y del mundo de la política pública para la colaboración internacional y el estudio de los procesos de CTI y el desarrollo. Constituye el capítulo regional de la Global Network for Economics of Learning, Innovation, and Competence Building Systems (Globelics), con un foco similar, pero de escala global.

El trabajo de LALICS se estructura en torno a tres ejes: la investigación, la formación y la cooperación. Este libro resulta de la articulación de esos tres ejes, en la medida en que combina esfuerzos internacionales para la traducción de los resultados de investigación en una herramienta de soporte a las actividades formativas. De la misma forma, el marco conceptual que guía el desarrollo de los capítulos es el enfoque de los sistemas nacionales de innovación (SNI), por su capacidad para abordar fenómenos complejos, históricos y dinámicos (Dutrénit y Sutz, 2014) su impacto en los procesos de política pública en la región (Crespi y Dutrénit, 2013) y la acumulación de conocimiento que se ha producido en el interior de la red desde cada uno de sus ejes y en diálogo con otras

redes. Un claro ejemplo son los libros organizados desde la Universidad Nacional de General Sarmiento en la Argentina, que han generado una serie de contenido útil para las actividades formativas en la región desde una visión propia, concebida por la comunidad académica latinoamericana con un foco claro en las siguientes generaciones; los libros *Repensando el desarrollo latinoamericano: una discusión desde los sistemas de innovación*, editado por Erbes y Suárez (2016), y *Teoría de la innovación: evolución, tendencias y desafíos*, editado por Suárez, Erbes y Barletta (2020), conforman el acervo base al cual esta obra se quiere sumar. En particular, nos interesa señalar dos tipos de contribuciones esperadas de este nuevo esfuerzo colectivo:

1. Conceptual y situada. Cada uno de los capítulos se inicia con una breve presentación del marco conceptual desde el cual se piensan los métodos y metodologías, que en general responde al enfoque de los SNI. Seguidamente, se explican detalladamente las formas en que esos métodos y metodologías pueden ser aplicados para el estudio de los procesos de CTI y finalmente se presenta un caso de aplicación práctica, desde el cual es posible reflexionar en torno a las fortalezas, debilidades, limitaciones y potencialidades de las técnicas para objetos de estudio situados en el contexto de los países latinoamericanos. Por ejemplo, hay un amplio espacio para considerar cómo hacer un mejor tratamiento de la heterogeneidad estructural en el sector productivo, particularmente frente a los supuestos de normalidad que están presentes en los métodos cuantitativos. Y, también, está la oportunidad de potenciar los espacios de aplicación de los métodos cualitativos, ampliando las posibilidades que pueden ofrecer herramientas distintas a los estudios de caso.
2. Formativa y crítica. Esta obra no es un punto de llegada, sino de partida, en dos direcciones. La más importante está alrededor de las personas para quienes se pensó: la idea es contribuir con la formación de estudiantes interesadas e interesados en los estudios de CTI, quienes recurrentemente hacen referencia a los desafíos que enfrentan cuando buscan desarrollar sus trabajos de investigación; desafíos que luego acompañan a la comunidad científica (incluso a quienes tienen más tiempo haciendo investigación) cuando se enfrentan a nuevos problemas. De ahí el segundo punto

de partida: esta obra pretende ser un inicio para generar un debate claro y enfocado sobre cómo realizamos los estudios de CTI y las implicaciones que esto puede tener en el proceso de investigación, tanto en términos de la agenda de investigación que desarrollamos como en sus resultados y en las consideraciones sociales que estos tienen. La propuesta es que sea una obra viva, que pueda evolucionar con ediciones futuras, compañera de los procesos de desarrollo en América Latina.

El origen de esta obra se encuentra en la Segunda Conferencia LALICS, realizada en 2018 en la Ciudad de México, en la que a partir de los debates e intercambios se identificó la necesidad de generar un espacio de reflexión, pero también de formación respecto de las metodologías deseables y posibles para abordar los problemas de investigación dentro del campo de los estudios de CTI. En ese año, y como continuidad de un proyecto editorial de mayor data, se lanzó una convocatoria abierta a toda la red para presentar postulaciones a capítulos para la obra en cuestión. Se recibieron alrededor de cuarenta y siete postulaciones, de las cuales veintiséis fueron seleccionadas para participar y se compilan en esta obra. Las contribuciones provienen de la Argentina, Brasil, Chile, Colombia, México, Perú y Uruguay, en algunos casos en colaboración con investigadores/as radicados/as en Alemania, España y Estados Unidos. La diversidad de problemáticas, la originalidad en los abordajes empíricos y la posibilidad de replicar experiencias de investigación dan cuenta de la dinámica de producción de conocimiento de nuestra red y la importancia de la cooperación regional para la superación de los desafíos del desarrollo.

Estructura: metodologías y métodos desde una perspectiva contextualizada

Como se mencionara, el objetivo es poner a disposición del público interesado un conjunto de métodos y diseños metodológicos que sirvan de referencia para el abordaje de diversos problemas de investigación dentro del campo de los estudios de CTI. De manera deliberada, hemos omitido la reflexión epistemológica respecto de los métodos y el análisis de los resultados, no porque no la creamos necesaria o relevante sino porque

hemos decidido enfocar los esfuerzos en la generación de material de consulta para estudiantes de posgrado, académicas y académicos.

Este libro se estructura en tres volúmenes, que contienen además dos secciones transversales a todos ellos. La división de los volúmenes responde a la tradicional separación entre el análisis cualitativo y las técnicas mixtas, en la que además se combinan abordajes emergentes en materia de medición. Esta segmentación tiene por objeto simplificar el acceso a los diferentes métodos, aunque en detrimento de la posibilidad de combinaciones entre diferentes metodologías. No obstante, esperamos que esta introducción y las secciones transversales a todos los volúmenes despierten el interés por nuevas combinaciones y estructuras de clasificación al momento de abordar un problema de investigación.

La sección transversal incluye dos contribuciones, una sobre democratización del conocimiento y otra sobre transversalización de la perspectiva de género. En "I. La perspectiva democratizadora en el análisis de los procesos sociales de investigación e innovación", Rodrigo Arocena y Judith Sutz reflexionan en torno a la construcción del objeto de estudio y su abordaje empírico, la demanda por sus resultados, la posibilidad efectiva de apropiación y los procesos de evaluación en relación con los procesos de democratización del conocimiento. En este marco, proponen reflexionar sobre la forma en que los procesos de investigación contribuyen con la expansión del poder al pueblo y la disminución de las desigualdades que emergen de una distribución inequitativa del conocimiento entre personas y naciones. En "II. Desafíos para la investigación. ¿Ciegos o con perspectiva de género?", Nora Goren aporta pautas para pensar la transversalización de la perspectiva de género en los procesos de investigación. Para ello, recorre con una perspectiva histórica y situada el concepto de género, el uso del lenguaje inclusivo y por qué resulta necesario contemplar las desigualdades sexogenéricas durante todo el proceso de creación de conocimiento.

Volumen 1. Métodos cualitativos

El volumen 1 compila siete capítulos centrados en los abordajes cualitativos de los procesos de CTI. En el capítulo 1 "Uso de diseños flexibles de investigación para el análisis de procesos de ciencia, tecnología, innovación y sociedad (CTIS) en América Latina", Elena Mendoza y Marcela Amaro analizan el uso de métodos flexibles en el estudio de problemas

de CTI en América Latina. Su trabajo aporta al debate respecto de dos estrategias metodológicas, los estudios de caso y los estudios etnográficos, los que son luego aplicados para el estudio de los procesos de transición hacia el desarrollo sostenible de comunidades locales de México.

En el capítulo 2 "Possibilidades e implicações da abordagem fsQCA para analisar dimensões qualitativas da inovação: os casos brasileiros de desenvolvimento tecnológico e institucional do setor produtivo de defesa e capacidade de absorção de firmas na interação universidade-empresa", Orlando Martinelli, Júlio Eduardo Rohenkohl y Janaína Ruffoni presentan el método de conjuntos difusos de análisis cualitativo comparativo (fsQCA por sus siglas en inglés) y proponen una forma de articulación epistémica integrada de todas las etapas de la investigación, incluida la reflexión respecto de variables lingüísticas y la traducción en forma de cuestionarios de la información a relevar. Aplican luego el análisis al caso de la promoción de un polo de defensa y la interacción universidad-empresas en el Estado de Rio Grande do Sul, Brasil.

El capítulo 3 "Una guía rápida para realizar investigación basada en la estrategia de estudios de caso: aplicaciones para el estudio de capacidades tecnológicas", Gabriela Dutrénit, Arturo Torres y Alexandre O. Vera-Cruz presentan los estudios de caso enfocados en análisis a nivel de empresa, como una herramienta para responder preguntas de investigación sobre cómo se desarrollan los procesos en el interior de ellas, así, plantean distintos tipos de estudios de caso y estrategias para su diseño.

En el capítulo 4, "ATLAS.ti: una herramienta para el desarrollo de métodos de análisis cualitativo", Soledad Rojas-Rajs y Marcela Suárez muestran cómo el uso de un software dedicado al análisis cualitativo puede ser operacionalizado en distintos proyectos de investigación. Además de presentar recomendaciones prácticas, las autoras buscan dar una guía de lineamientos previos al uso de la herramienta que permita obtener resultados de investigación más rigurosos.

El capítulo 5 "Innovación y desarrollo desde un enfoque sistémico y contextualizado: el enfoque de los arreglos productivos locales en Brasil" escrito por José Cassiolato, Helena Lastres, Marcelo Matos y Micaela Mezzadra propone un abordaje metodológico para los estudios basados en el concepto ya clásico de arreglos productivos locales (APL) en los que se incluye la reflexión respecto de los criterios de selección, la delimitación del APL, la definición de actores, actoras e instituciones a relevar, los instrumentos de recolección de información y su posterior tratamiento. Su observación práctica se basa en el trabajo realizado desde la RedeSist

y la aplicación del enfoque de los APL al estudio de múltiples casos en Brasil y otros países de la región.

El capítulo 6, "El estudio de caso múltiple para el análisis de las proximidades en entramados locales de América Latina", está escrito por María Eugenia Castelao Caruana, Mariel de Vita y Pablo Lavarello, quienes analizan los aportes y limitaciones del método de casos múltiples en articulación con el marco de proximidades y las consideraciones necesarias para su aplicación a las condiciones de los países de América Latina. En el capítulo se sugiere adoptar un enfoque dinámico de las proximidades que incluya la tensión entre espacios nacionales y grupos económicos internacionales, el papel del Estado en los procesos de aprendizaje e innovación y las jerarquías de las proximidades a lo largo de estos procesos.

El volumen 1 concluye en el capítulo 7, "Propuesta metodológica desde el enfoque de sistemas sociotécnicos: el caso de la biotecnología aplicada a la salud en México", a cargo de Luis Jiménez, José Miguel Natera y Daniel Villavicencio. En este texto se muestra cómo operacionalizar el estudio integrado de sistemas sociotécnicos, mostrando alternativas para la sistematización de datos que permitan la caracterización del paisaje sociotécnico, el régimen sociotécnico y los nichos; utilizando el caso de la biotecnología con aplicaciones en salud en México.

Volumen 2. Métodos cuantitativos

En el volumen 2 se incluyen once capítulos que abordan diferentes técnicas cuantitativas para el estudio de los procesos de CTI. En el capítulo 1 "La aplicación de técnicas de análisis multidimensional para el estudio de la organización del trabajo", Sonia Roitter y Analía Erbes sistematizan experiencias prácticas de aplicación del análisis factorial de correspondencias múltiples (AFCM) y de clúster en un marco de análisis multidimensional. Aplican luego este bagaje metodológico al caso de la relación entre la organización del trabajo y los procesos de innovación entre empresas argentinas.

En el capítulo 2, "Consideraciones para realizar estudios comparados utilizando datos provenientes de las encuestas de innovación en América Latina y el Caribe", Sandra Zárate, Nadia Albis, José Miguel Natera, Erika Sánchez y Fernando Vargas realizan un balance sobre las principales implicaciones metodológicas al momento de realizar estudios comparativos basados en las encuestas de innovación realizadas en la región. El

capítulo compila la revisión de doce encuestas de innovación aplicadas durante la ventana de observación de cinco años (2012-2016).

El capítulo 3 "Desafíos para los países de América Latina y el Caribe en la medición de la innovación frente a la edición 2018 del *Manual de Oslo*", escrito por Mónica Salazar, Nadia Albis, Sandra Zárate y Fernando Vargas avanza en la discusión iniciada en el capítulo anterior, al realizar el análisis y los desafíos que enfrentan los países de la región al momento de implementar las nuevas directrices brindadas en el *Manual de Oslo* en su versión 2018 en sus encuestas de innovación. En el capítulo se realiza una comparación entre ediciones y se revisan los conceptos y definiciones principales, las categorías de medición y análisis propuestas, las metodologías sugeridas y los métodos de recolección de información.

En el capítulo 4 "¿Complementariedad o sustitución sobre el comportamiento innovador? Desarrollo de la estrategia empírica a partir del test de supermodularidad", Carlos Bianchi y Pablo Blanchard desarrollan estrategias empíricas para estimar efectos de supermodularidad (complementariedad) o submodularidad (sustitución) de determinados eventos sobre el comportamiento innovador de las empresas. Los autores presentan ejemplos aplicados, los fundamentos lógico-matemáticos y las rutinas de programación para los software Stata y R basados en encuestas de innovación, y las consideraciones necesarias al momento de analizar empresas localizadas en países latinoamericanos y sus respectivas encuestas.

El capítulo 5 "Modelado con ecuaciones estructurales: una herramienta para observar y relacionar lo inobservable", de Soledad Contreras y Natalia Gras consiste en el estudio de las posibilidades que ofrecen los modelos de ecuaciones estructurales en el campo de CTI a partir de dos aplicaciones. Por un lado, el estudio sobre las relaciones entre dimensiones asociadas a la evaluación académica y los modos de producción de conocimiento en México. Por el otro, el desarrollo de un índice de vulnerabilidad energética para Montevideo. En ambos casos, las autoras presentan el conjunto de procedimientos implementados para desarrollar y validar ambos modelos y reflexionan en torno a la interpretación de los resultados y las limitaciones del modelo.

En el capítulo 6 "Uso de paneles de datos para la evaluación de la política de innovación", Florencia Fiorentin, Mariano Pereira y Diana Suárez reflexionan en torno a la evaluación de los procesos de asignación e impacto de la política de innovación a nivel de la firma a través de la combinación de estrategias metodológicas vinculadas con los estudios

dinámicos en paneles de datos. Además del desarrollo de los modelos, en el capítulo se discuten las implicancias del marco conceptual de la política al momento de evaluarla, y se reflexiona en torno a las especificidades para América Latina y el Caribe y se proveen algunas recomendaciones generales e implicancias para el análisis.

En el capítulo 7 "Regresiones transversales. ¿Tienen pertinencia en el estudio de la innovación?", Andres Felipe Zambrano-Curcio, Norida Constanza Vanegas Chinchilla, Nicolas Fuentes y Jana Schmutzler analizan el método de regresiones transversales y cómo –a pesar de su desventaja de no poder establecer relaciones de causalidad entre variables– constituyen potentes herramientas de base para investigaciones cuantitativas. Además de revisar la literatura basada en el estudio de encuestas de innovación, el capítulo presenta distintos modelos de regresiones a partir de datos discretos, reflexiona en torno a las problemáticas que surgen al momento de la aplicación y el análisis y propone algunas soluciones.

El capítulo 8 "Métodos de análisis de insumo producto: aplicaciones a la CTI en América Latina", Patieene Alves-Passoni, Leobardo Enríquez, Rosa Gómez Tovar, Brenda Murillo-Villanueva y Martín Puchet Anyul (coord.) presentan una serie de ejercicios que muestran la repercusión de las actividades de ciencia, tecnología e innovación (CTI) en la estructura mesoeconómica de los países o de conjuntos de ellos utilizando el enfoque de insumo, producto y su metodología; abordando temas como la deflactación, la descomposición y las posibilidades de usar la matriz interpaíses junto con una serie de reflexiones finales sobre los usos y posibles extensiones del conjunto de métodos asociados.

En el capítulo 9 "Cointegración para el estudio de la evolución de los sistemas de innovación", José Miguel Natera y José Ignacio Ponce presentan el método de cointegración como una alternativa para incorporar la dimensión temporal de forma estructurante en el análisis cuantitativo de los sistemas de innovación, particularmente en su visión nacional y en la relación que estos tienen con los procesos de desarrollo económico. Así, presentan tres alternativas para el uso del método (análisis de regresiones de cointegración, cointegración en series temporales y cointegración en panel), acompañadas de una discusión sobre los datos necesarios para su aplicación y de posibles extensiones en su uso.

El capítulo 10 "Análisis multinivel: retos y oportunidades para el estudio de los procesos de innovación en América Latina", escrito por Guillermo Orjuela-Ramírez y Julio César Zuluaga presentan las principales ventajas y oportunidades de investigación basadas en los análisis con

un enfoque multinivel, considerando cómo aplicar la técnica multinivel mediante un ejercicio empírico usando datos de encuestas de innovación disponibles en América Latina, reflexionando sobre las ventajas, desventajas y limitaciones.

En el capítulo 11 "Drivers tecnológicos del crecimiento. Indicadores agregados y el tratamiento agregado de la heterogeneidad", Nuria E. Laguna Molina y Ana Urraca Ruiz proponen una reflexión respecto de la identificación de los principales conductores tecnológicos del crecimiento (drivers) y su naturaleza para avanzar luego en una propuesta de medición y análisis cualitativo a través de un modelo de crecimiento con datos de panel.

Volumen 3. Métodos mixtos y emergentes

El volumen 3 compila ocho capítulos en los que hemos incluido técnicas mixtas y emergentes. Algunas de esas técnicas presentan fuertes solapamientos y articulaciones con los capítulos de los volúmenes anteriores, pero entendemos que su tratamiento conceptual por separado contribuye a identificar fortalezas, debilidades y especificidades de cada una de las técnicas cuando se aplican a casos localizados en los países de América Latina. Así, en el capítulo 1 "Análisis del proceso de innovación en América Latina a partir de la combinación metodológica de la modelación basada en agentes y el análisis de redes sociales" Walter Lugo Ruiz-Castañeda, Juan F. Franco-Bermúdez y María Luisa Villalba-Morales presentan una propuesta para combinar la modelación basada en agentes con el análisis de redes sociales y analizan algunos trabajos que han combinado las metodologías en América Latina.

En el capítulo 2 "El análisis de redes sociales: una herramienta de análisis para entender los patrones de creación y difusión de conocimiento. Su aplicación a colaboraciones científicas y de conocimiento interorganizacionales", Lilia Stubrin y Cecilia Tomassini exploran la metodología del análisis de redes sociales (ARS) como herramienta para entender los patrones de creación y difusión de conocimiento. Tal como se expresa en su título, el foco del capítulo está puesto en la aplicación del ARS al análisis de las colaboraciones científicas e interorganizacionales. Las autoras proponen una reflexión en torno a la técnica y sus implicancias y limitaciones para estudiar fenómenos en Latinoamérica.

El capítulo 3 "Aplicaciones de la teoría de grafos al análisis de sistemas de innovación y espacios tecnológicos", escrito por Ana Urraca Ruiz, Pedro Miranda y Vanessa de Lima Avanci, se presentan las nociones básicas de teoría de grafos y cómo se pueden construir grafos para representar y analizar bases de conocimiento. En el capítulo se analizan los conceptos clave para la aplicación de la técnica, se revisan las principales fuentes de información disponibles y se proponen indicadores para analizar sus propiedades.

En el capítulo 4 "Modelado y simulación de problemas de CTI con dinámica de sistemas", Mauricio Uriona Maldonado y Milton M. Herrera presentan el método de simulación de dinámica de sistemas y su aplicación en problemas de CTI en América Latina. El capítulo incluye una descripción de los conceptos y premisas importantes para la aplicación del método, las etapas del proceso de modelado, junto con ejemplos de aplicación para cada una, y una reflexión respecto de las potencialidades y limitaciones para su aplicación en casos localizados en América Latina.

En el capítulo 5 "El método de revisión de la literatura estructurada para los estudios de CTI", Caroline Rodrigues Vaz y Mauricio Uriona Maldonado describen el método de revisión estructurada de literatura, sus pasos y etapas y, para cada una de las etapas, se proveen ejemplos de aplicación en un tema relacionado con los procesos de CTI, conjunto de software informático necesario y las implicancias para América Latina y el Caribe.

El capítulo 6 "Modelación y simulación como herramientas para la comprensión de fenómenos emergentes en la difusión y transferencia de tecnologías", escrito por William Alejandro Orjuela Garzón y Santiago Quintero Ramírez reflexionan en torno a las diferentes aplicaciones de la modelación basada en agentes (MBA) con énfasis en los procesos de transferencia y difusión de tecnologías. En el capítulo se parte de un análisis respecto de los paradigmas de simulación y se presenta el proceso metodológico para construir, verificar y validar un MBA, así como también las herramientas computacionales para su construcción.

El capítulo 7 "El análisis semántico-estadístico como estrategia de abordaje metodológico: reflexiones sobre su pertinencia en el estudio de problemáticas latinoamericanas" de Matías Milia y Rodrigo Kataishi consiste en la presentación y análisis de técnicas para entender y procesar datos textuales en el estudio de las dinámicas de CTI. Se propone una reflexión en tres dimensiones: conceptual, analítica e informática, a partir de las cuales se avanza en el análisis del impacto en las prácticas de investigación y el potencial de su aplicación en América Latina.

El capítulo incluye además la presentación de cuatro casos de estudios empíricos a partir de los cuales se desarrollan los conceptos clave.

Finalmente, en el capítulo 8 "Modelos estructurales cualitativos para el estudio y comprensión de los procesos de ciencia tecnología e innovación", Mayela Saraí, López-Castro, Nayely Martínez, Natalia Gras y José Miguel Natera muestran cómo los modelos estructurales cualitativos son una herramienta para deducir e interpretar estructuras, dimensionalidad y relaciones subyacentes en los fenómenos complejos. Utilizan información cualitativa y para ello presentan dos métodos: el modelado estructural interpretativo total y el análisis estructural-causal cualitativo, señalando los pasos a seguir para su aplicación.

Así, a través de veintiséis capítulos y dos secciones especiales, esta obra compila un amplio abanico de estrategias metodológicas para el estudio de los procesos de CTI en la región latinoamericana. Se trata, desde luego, de una selección arbitraria de temas pero que por su proceso de selección y los debates que tuvieron lugar en el marco de este proyecto, reflejan en gran medida muchos de los temas de investigación que están siendo objeto de debate en la región y en el mundo. Esperamos con ello contribuir a ese debate, con una mirada desde y para América Latina.

Bibliografía

Cohen, W. M. y Levinthal, D. A. (1989). "Innovation and Learning: The Two Faces of R&D". *Economic Journal*, vol. 99, nº 397, pp. 569-596. DOI: https://doi.org/10.2307/2233763.

Crespi, G. y Dutrénit, G. (eds.) (2013). *Políticas de Ciencia, Tecnología e Innovación para el desarrollo: la experiencia latinoamericana*. México: FCCyT/LALICS.

Dutrénit, G. y Sutz, J. (eds.) (2014). *Sistemas de Innovación para un Desarrollo Inclusivo. La experiencia latinoamericana*. México: Foro Consultivo Científico y Tecnológico.

Erbes, A. y Suárez, D. (comps.) (2016). *Repensando el desarrollo latinoamericano: Una discusión desde los sistemas de innovación*. Los Polvorines: UNGS. Disponible en: https://repositorio.ungs.edu.ar/bitstream/handle/UNGS/275/712_RepensarDesarollo_WEB.pdf?sequence=1&isAllowed=y.

Suárez, D.; Erbes, A. y Barletta, F. (comps.) (2020). *Teoría de la innovación: evolución, tendencias y desafíos*. Los Polvorines/Madrid: UNGS/ Ediciones Complutense.

I.
La perspectiva democratizadora en el análisis de los procesos sociales de investigación e innovación

Rodrigo Arocena y Judith Sutz

Presentación

El conocimiento es poder; la vieja frase se refiere a una realidad aún más antigua, pero debe ser reformulada y afinada: el conocimiento científico y tecnológico de punta es hoy una fuente de poder hoy más que nunca. La afirmación, como debiera ser evidente, no tiene carácter normativo sino descriptivo: guste o no guste, así parecen ser las cosas.

El poder de un grupo puede ser considerado como la posibilidad que tiene de lograr sus fines mediante el control de su entorno natural y social. Tecnología y organización son medios fundamentales de poder. La expansión del conocimiento científico ha alimentado el gravitante despliegue de las tecnologías de la producción, la destrucción, la comunicación, la información, etc. Las organizaciones más fuertes –empresas, Estados, ejércitos, iglesias, etc.– suelen tener posibilidades grandes de usar las tecnologías para robustecer su poder. Esto sucede "hacia afuera", como dominio de la organización sobre su entorno, y también "hacia adentro", como dominio en el interior de la organización de quienes dirigen y controlan sobre los dirigidos y controlados.

La creación y el uso de conocimiento son actividades que involucran a la vez la cooperación y el conflicto de variados actores. En los procesos sociales de investigación e innovación hay ganadores y perdedores. Ignorarlo no ayuda a los muchos perdedores y hace difícil enfrentar las olas

"anticiencia" con sus numerosos perjuicios, particularmente su notable contribución a la degradación ambiental y climática. Ignorarlo tampoco ayuda a comprender que nunca hay una única forma de encarar y solucionar un problema dado, que lo que se presenta como ineluctabilidad tecnológica –"no hay alternativa"–, derivada de una superioridad medida sobre todo en términos técnicos, no es tal si se toman en cuenta otros parámetros.

En general, los impactos diferentes del conocimiento avanzado en las condiciones de vida de distintos grupos humanos, incluyendo la expansión o la disminución de la desigualdad entre ellos, dependen en gran medida de las relaciones de poder: prestarle a tal cuestión una atención especialísima es la principal sugerencia metodológica que orienta las páginas siguientes. Las relaciones de poder se expresan, por ejemplo, en la conformación de las agendas de investigación y de innovación: ¿los problemas de quiénes se toman en cuenta? También se expresan en las heurísticas seguidas para abordar problemas: las soluciones encontradas podrán ser soluciones para algunos y no para otros. En ocasiones, las relaciones de poder detrás de qué se investiga y qué no, en torno a qué se innova y qué no, aparecen de forma inmediata; en otras, se vuelven mucho menos visibles justamente por no reconocer que en los procesos sociales de investigación e innovación hay ganadores y perdedores.

Este manual será sin duda un aporte sustantivo para elevar el nivel de los estudios sobre ciencia, tecnología, innovación y sociedad en América Latina. La invitación a contribuir al mismo nos honra y, además, nos ofrece una oportunidad para subrayar que tales estudios se benefician cuando aportan una pluralidad de enfoques en combinación con el diálogo entre aproximaciones distintas. En un trabajo anterior sugerimos vías para ello a través de la construcción de una agenda compartida (Arocena y Sutz, 2020). Aquí, de manera harto sintética, pero esperemos que autocontenida, se argumenta que al mismo objetivo puede colaborar el que distintos análisis tengan en cuenta una perspectiva transversal sobre la democratización del conocimiento. La misma, en definitiva, no es sino una parte de la respuesta a la pregunta, también muy vieja, acerca de por qué y para qué impulsar la investigación y la innovación.

Actores y poder en los sistemas de innovación

Las concepciones más elaboradas de los sistemas nacionales de innovación (SNI) atienden prioritariamente las tecnologías, instituciones e influencias mutuas entre unas y otras. Consideran a la innovación como proceso "distribuido" e interactivo, lo que implica que variados actores participan, que las iniciativas pueden surgir en ámbitos diferentes y que los resultados dependen altamente de los vínculos entre los participantes. Para tales concepciones la innovación es más efectiva –genera mayor poder, es decir, incrementa la posibilidad de que algunos logren sus fines mediante el control de su entorno natural y social, diríamos por cuenta nuestra– cuanto más sistémica sea, vale decir, cuanto más sólidas y estables sean las interacciones entre los actores involucrados.

Ahora bien, ¿quiénes orientan la innovación y se benefician de ella? Más en general, ¿cómo inciden las modalidades prevalecientes de la innovación en la expansión del poder de la nación y en su distribución interna? Para encarar estas cuestiones cabe considerar el triángulo de Sabato como un triángulo nuclear del poder en un SNI dado: el vértice estatal concentra el poder político y militar, el vértice productivo lo hace con el poder económico, y el tercer vértice es el principal generador de conocimientos.

El desarrollo económico acelerado de Corea del Sur y Taiwán llevó a ambos países a superar la condición periférica; fue comandado por la alianza entre las cúpulas del Estado y el empresariado, forjado por dictaduras militares y cementado por la ideología del nacionalismo tecnológico. Este último les permitió enfrentar con éxito el poder ideológico, acompañado por el poder económico, de los intereses extranjeros que les indicaban cuál era su lugar en la división internacional del trabajo, lo que promovió la generación nacional de conocimiento avanzado y su incorporación a la producción. El triángulo nuclear del SNI se afianzó; el poder de cada nación se expandió y su distribución interna tuvo lugar durante largo tiempo en contextos antidemocráticos y en desmedro de los trabajadores.

Una experiencia anterior de superación de la condición periférica fue protagonizada por los países escandinavos; tuvo como columna fundamental la conformación de SNI mucho más inclusivos que en el Este de Asia, basados en un entendimiento de largo plazo entre gobiernos democráticos, empresariado y sindicatos, así como en el respaldo a la investigación científica, de modo tal que la ideología del Estado de bienestar

incidió significativamente en la orientación de la innovación y en la diversidad de actores que en ella participaron. El desarrollo económico consiguiente fortaleció el poder de cada uno de esos países mientras que la distribución interna del poder ha sido de las menos desiguales que la historia registre.

En América Latina, sin desmedro de su enorme diversidad, cabe decir que los SNI han sido por lo general más virtuales que reales. Las dinámicas prevalecientes de la economía han generado comparativamente escasas demandas de conocimiento avanzado y las han dirigido mayoritariamente hacia el exterior; los gobiernos no han logrado revertir ese fenómeno central y en la mayor parte de los casos no lo han intentado. Así, el conocimiento endógenamente generado ha sido fuente de poco poder de cada nación como tal y en el interior de ella, por lo cual los actores sociales y políticos no le han prestado demasiada atención; en tal contexto, la condición periférica ha seguido vigente. Ella tiene entre sus cimientos la arraigada dependencia ideológica, que privilegia la investigación y la innovación externas por el solo hecho de serlo, y favorece la desprotección frecuente de los procesos internos de aprendizaje cognitivo y productivo.

Las consideraciones precedentes sugieren que, a la hora de analizar los procesos sociales de investigación e innovación, se preste especial atención a establecer cuáles son los actores que realmente forman parte del SNI y en qué medida inciden en sus orientaciones. Al respecto conviene no olvidar que existen actores fuertes en conocimiento y actores débiles en conocimiento. Empresas grandes con departamentos de I+D formalmente constituidos ejemplifican el primer caso, empresas pequeñas sin siquiera un técnico de nivel universitario, el segundo; las políticas dirigidas a la producción no pueden ser de talla única y los analistas no deben ignorarlo. En el mundo periférico, la mayor parte de los grupos y actores son débiles en conocimiento, lo que de hecho los margina de los sistemas de innovación y refuerza el carácter subordinado que a menudo les imponen las relaciones sociales prevalecientes. Políticas de conocimiento y de innovación diseñadas en el olvido de este hecho central refuerzan la subordinación.

Lo que aquí se plantea se apoya en la afirmación normativa de que el conocimiento debe respaldar el desarrollo humano sustentable entendido como la expansión de las capacidades y libertades de la gente para vivir vidas que se consideran valiosas y en formas que tiendan a proteger y reparar el ambiente más que a deteriorarlo. No se trata de

discutir aquí acerca de los variados orígenes del conocimiento: se trata sí de reconocer y tomar centralmente en cuenta que, como se decía al comienzo, el conocimiento científico y tecnológico de punta es hoy una fuente de poder más que nunca en la historia. Por eso mismo cobra importancia la afirmación normativa ligando conocimiento con desarrollo humano sustentable.

Tiene carácter tanto normativo como fáctico la afirmación de que considerar a la gente no como pacientes sino como agentes es imprescindible para promover ese tipo de desarrollo. En todas partes se multiplican los esfuerzos e iniciativas que enfrentan a la falta de sustentabilidad, a las variadas manifestaciones de la desigualdad y al avance de los autoritarismos. Es una comprobación fáctica que esas formas de la agencia no pueden tener éxito profundo y a largo plazo si no se insertan en redes y sistemas fuertes en conocimiento. El desarrollo requiere pues que la democratización, como expansión del poder del pueblo, alcance al conocimiento, lo que significa tanto fortalecer cognitivamente a los países periféricos como disminuir en todos los países las desigualdades basadas en el conocimiento. Esta perspectiva general para el análisis de los procesos sociales de investigación e innovación lleva a sugerencias concretas como las que se mencionan a continuación, que no pretenden ser realmente guías metodológicas, sino tan solo aportes para la reflexión.

Algunas dimensiones del análisis

Las *agendas* predominantes en la investigación y la innovación constituyen un indicador sustancial, y en buena medida accesible al análisis, del condicionamiento social de ambas actividades, especialmente de los intereses y las ideas que mayor gravitación tienen en ellas. Los problemas que se priorizan o postergan anticipan en buena medida quiénes se beneficiarán y, también, quiénes se perjudicarán con las nuevas técnicas que se pongan a punto. Dos de los ganadores del Premio Nobel de Economía 2019 dicen que, en materia de inteligencia artificial vinculada a la salud, se privilegia la automatización de los procedimientos de las compañías de seguros y se desatiende la elaboración de procedimientos que ayuden a las personas recién operadas a recuperarse en sus hogares con ayuda de personal especialmente preparado (Banerjee y Duflo, 2019). Buena parte de lo que importa queda en evidencia mediante este ejemplo: la atención a los intereses de pocos en desmedro de los beneficios potenciales para

muchos, la destrucción de empleo en lugar de la creación de tareas calificadas, las posibilidades que se abren para las políticas públicas con el fin de reorientar esfuerzos fundamentales.

La dimensión recién destacada se vincula estrechamente con la que tiene que ver con las *demandas* de conocimiento nuevo. Desde este punto de vista, en principio, la que más gravita es la demanda solvente, vale decir, la respaldada por el poder de compra. Es también la menos difícil de registrar; pero no debiera ser la única a considerar. Tan conocida como sintéticamente elocuente es la afirmación de la Organización Mundial de la Salud según la cual el 90% de los recursos destinados a la investigación en este campo se vincula con los problemas que afligen al 10% de la población mundial. Así, se pone de manifiesto la existencia de una demanda social de conocimientos cuya mejor atención multiplicaría los beneficios de la investigación y la innovación, cuestión por ende fundamental para las políticas en este campo. En los análisis de esta temática vale la pena tener en cuenta que la demanda efectiva puede ser bastante menor que la demanda potencial no solo por falta de fondos, sino por cuestiones propiamente cognitivas; en efecto, las necesidades de conocimiento nuevo, aun si están respaldadas por un poder de compra, no siempre son fáciles de traducir a problemas formulados de modo que puedan incorporarse a la agenda de investigación e innovación.

Las agendas y demandas tienen mucho que ver con la desigualdad ligada al conocimiento. Ambas se vinculan directamente con la dimensión de las *heurísticas* de la investigación y la innovación. ¿Qué soluciones se buscan, mediante cuáles estrategias cognitivas, contando con cuántos recursos, considerando satisfactorios qué tipo de resultados? Esta es una temática inmensa de la cual aquí se mencionará una faceta, la capacidad de innovar en condiciones de escasez que ha forjado en varias regiones del Sur global heurísticas propias a partir de talentosas respuestas originales a la penuria de recursos (Srinivas y Sutz, 2008). La pandemia ha multiplicado particularmente en América Latina los ejemplos de ese tipo de innovación frugal basada en investigación del más alto nivel, cuyo balance de costos y beneficios ha despertado el interés del Norte. Heurísticas semejantes pueden contribuir, junto con el imprescindible incremento de la inversión en investigación e innovación, a enriquecer sus agendas y a prestar mayor atención a la demanda social. Ambas son formas concretas de democratizar el conocimiento. Las heurísticas predominantes han sido generadas en el Norte, por lo general en contextos de relativa abundancia de recursos materiales, lo cual no contribuye al uso frugal

de los bienes naturales. Esta dimensión se vincula pues estrechamente con la problemática de la sustentabilidad.

La dimensión de los *objetivos* generales de la investigación y la innovación merece también ser considerada explícitamente. Durante largo tiempo se asumió con poca discusión que el propósito central de la innovación técnico-productiva es contribuir a la competitividad económica. Aun dentro de este marco, los requisitos de conocimiento pueden variar considerablemente en profundidad y amplitud cuando los objetivos se especifican algo más; para ponerlo de manifiesto basta recordar la distinción clásica destacada por la CEPAL entre competitividad "espuria" y competitividad "auténtica". La primera se basa particularmente en los bajos salarios y en el uso sin contemplaciones de los recursos naturales, mientras que la segunda pretende ser compatible con el desarrollo humano sustentable. Agendas, demandas y heurísticas pueden ser distintas en uno u otro caso, y diferir todavía mucho más si se apunta a objetivos que no se reduzcan al fomento de la competitividad. Un ejemplo de ello lo constituyen los programas de investigación e innovación orientados a respaldar la inclusión social, que trabajosamente se han venido abriendo paso en tiempos recientes.

En la perspectiva que aquí se esboza, el análisis de los procesos de generación y uso del conocimiento debe también prestar atención a las *disciplinas involucradas*. Los problemas de la realidad no vienen por lo general encuadrados por disciplinas, pero las dificultades del trabajo interdisciplinario suelen ser grandes y más bien inmunes a discursos en pro del holismo. El conocimiento avanzado exige especialización, pero, cuando se apunta a usarlo eficazmente, rara vez es asunto de una sola especialidad. Menos lo es cuando se quiere comunicar bien lo que se hace. Y mucho menos cuando se quiere tener en cuenta las interacciones entre ciencia, tecnología y sociedad. En suma, no conviene descuidar la dimensión interdisciplinaria.

Igualmente digna de atención es la dimensión de los *actores* que se consideran. El enfoque que aquí se sugiere puede ilustrarse a partir de la cuestión de las relaciones entre academia y producción. Su frecuente denominación como "universidades y empresas" hace obvio el descuido de una gama de actores –sindicatos, cooperativas, asociaciones de pequeños productores, etc.– lo que, por un lado, implica perder sus potenciales aportes a los procesos innovativos y, por otro, dada la orientación predominante de las agendas de innovación, puede dejarlos al margen de sus resultados. El grado de compromiso con la democratización puede estimarse a partir

de lo que se hace en pro del involucramiento real de ese tipo de actores en los procesos de investigación e innovación que les conciernen.

Actores distintos tienen tipos diferentes de conocimiento valiosos pero que pueden o no ser tenidos en cuenta. Se aprende desde lo que se sabe, y los aprendizajes en general expanden las capacidades para desempeñarse como agentes. La concepción de la innovación como proceso distribuido e interactivo recuerda que hay que prestar atención a lo que se puede hacer en variados ámbitos y a partir de la combinación de aportes diversos. Lo recién anotado destaca la dimensión de los *procesos interactivos de aprendizaje*, en los que todos los actores involucrados pueden aprender de los demás en la búsqueda conjunta de soluciones a problemas significativos; detectar y analizar procesos semejantes ayuda a captar en qué medida la innovación, en un lugar geográfico o un sector de actividad, tiene carácter sistémico.

Gran parte de lo antedicho tiene que ver con el tipo de innovación que se tiene en cuenta o, por el contrario, se deja fuera de los estudios, aunque sea implícitamente. Esto último suele suceder con el tipo de innovación que tiene carácter más o menos informal, lo que en modo alguno equivale a trivial. Ello tiene particular importancia para una perspectiva pensada "desde y para Latinoamérica", como lo reivindica este manual. En efecto, en el Sur en especial, la innovación formal está lejos de ser la única digna de atención. Lo que se incluye o no en el *registro* de actividades creativas es una dimensión del análisis que mucho puede decir acerca de los sistemas de innovación, sus componentes, sus capacidades y su potencial de futuro.

Por este camino, como por tantos otros, se llega a la dimensión de la *evaluación*: ¿con qué criterios se aprecian las actividades de investigación e innovación? Los predominantes son bastante pobres, privilegian indicadores más formalizados que sustantivos, subordinan lo cualitativo a lo cuantitativo y dejan de hecho en manos ajenas a la región la valoración de lo que en ella se hace. Si tales criterios son adoptados de manera acrítica y estereotipada, como si fueran los obvios, no se está analizando "desde y para Latinoamérica". Cosa no menos grave, se arriesga desatender investigaciones altamente originales e innovaciones socialmente muy valiosas.

En fin, la *apropiación* de los resultados de las labores creativas es una dimensión del análisis cuya importancia en esta perspectiva apenas precisa ser justificada. Incluye la consideración de las justificaciones, fundadas o ficticias, para la distribución de los beneficios ligados a la generación y uso de conocimientos. Tiene que ver con la mayor o menor

validez del sistema de patentes y con la conveniencia de reemplazarlas en cierta medida por otras recompensas, como los premios. En tiempos de pandemia, cuando los conflictos por el acceso a las vacunas ocupan los primeros lugares de la información, la cuestión hace evidente lo que significa la democratización del conocimiento.

Una hora latinoamericana

Mirando al continente desde nuestro rincón, parece que la creación endógena de conocimientos está haciendo aportes al enfrentamiento al covid-19 que la realzan como nunca antes a los ojos de la ciudadanía. Esta legitimidad configura una notable oportunidad; los estudiosos pueden contribuir a que no sea desaprovechada.

Las sugerencias consideradas en la sección precedente tienen como sustento común una afirmación central, que aquí corresponde explicitar: cada uno de nuestros países necesita y puede tener investigación e innovación nacional, de nivel internacional, con vocación social. Esos tres rasgos caracterizan a la creación de conocimientos orientada a su democratización. Corresponde analizar su vigencia en la realidad. Para ello puede no ser superfluo comentar brevemente sus respectivos contenidos. Se verá así que las dimensiones del análisis que se plantearon antes son algunas entre varias maneras posibles de "operacionalizar" dicha afirmación central.

Investigación e innovación nacional significan lo contrario tanto de parroquial o aislada como de subordinada o auxiliar. O'Donnell (2004) dice que en el Norte se espera que la labor académica del Sur consista en ayudar con datos primarios a la elaboración teórica que allí se hace y que aquí recibiremos como consumidores de productos terminados, en modo afín a la división del trabajo propia del sistema centro-periferia. La creación nacional debe apuntar a intervenir como protagonistas en los debates globales sobre ideas y alternativas, contribuyendo así a la autonomía cultural y política de los países de nuestra región latinoamericana.

La investigación e innovación de nivel internacional son necesarias porque nada menos que eso puede conformarnos cuando se trata de la creación de cultura y de las condiciones de vida en nuestros países. Que ese nivel no sea espurio e impuesto desde afuera sino auténtico y establecido en diálogos horizontales depende en gran medida de que cultivemos nuestras propias capacidades, que las conozcamos y confiemos en ellas.

La investigación e innovación con vocación social es lo que contribuye a que el poder del conocimiento no favorezca ante todo a los más privilegiados, como es la dinámica prevaleciente en nuestra época, sino que contribuya al bienestar material y espiritual de todos priorizando a los sectores más postergados.

Como ha vuelto a comprobarse durante la pandemia, es un dato de la realidad que en nuestra región las universidades públicas constituyen la principal sede de creación de conocimiento avanzado. Ellas tienen una gran responsabilidad en esta hora latinoamericana dramática y a la vez propicia para hacer de la investigación y la innovación palancas potentes del desarrollo humano sustentable. Estarán a la altura de tamaña oportunidad en la medida en que, en las antípodas de la autarquía academicista, multipliquen las colaboraciones con los actores reales o potenciales de los sistemas de innovación inclusivos y sustentables, incluyendo otros centros de generación de ciencia y tecnología. Esta cuestión es digna de la mayor atención en los análisis concebidos "desde y para Latinoamérica".

El análisis del análisis

Este manual discute métodos para analizar los procesos sociales de investigación (científica y tecnológica) y de innovación. En otras palabras, se ocupa de la generación de conocimiento acerca de la generación y uso del conocimiento. Tiene pues que ver con lo que se ha denominado "ciencia de la ciencia", actividad a la que corresponde aplicar la exigencia de "reflexividad" (Bourdieu, 2001), o sea de revisión crítica de la propia labor, que debe caracterizar a la ciencia en general. Ello lleva a sugerir que la perspectiva de la democratización del conocimiento debe figurar en dos instancias o niveles: primero, en los procesos que se analizan; segundo, en los propios análisis.

La democratización planteada tiene que ver con la índole del conocimiento científico, que no es ni uno entre otros sino una fuente superlativa de poder, ni tampoco la verdad, con mayúscula y definitiva. Esto último es bastante conocido, pero resulta a menudo disimulado para fortalecer las posiciones de quienes aparecen como voceros autorizados de la ciencia.

¿Qué puede significar democratizar el conocimiento acerca de la generación y uso del conocimiento? En principio, cabe poner en juego los mismos criterios planteados para la democratización del conocimiento en general. Por ejemplo:

- Involucrar en el estudio de los procesos de investigación e innovación a sus principales protagonistas, reales e incluso potenciales. Así cabe esperar resultados más ricos del análisis –democratizar el conocimiento incluye crear mejor conocimiento–y aplicaciones más fecundas de este. Lo sugerido se inscribe dentro del principio general de valorización de la agencia.
- Incluir en la presentación de resultados una parte autocontenida cuya comprensión no esté limitada a los especialistas y se extienda más allá de los hacedores de políticas. Esto parece fundamental para llegar a los actores débiles en conocimiento y para encarar el difícil problema de la lejanía de gran parte de los sectores populares respecto al conocimiento avanzado.

Recapitulación: políticas y política

Los análisis de los procesos sociales de investigación científica y tecnológica y de innovación apuntan, en medida considerable, a formular propuestas para las políticas públicas en ciencia, tecnología e innovación (CTI). No es un secreto que buena parte de tales propuestas son desatendidas. Esta cuestión no debiera ser obviada en dichos análisis, en la medida en que con ellos se procura contribuir a incrementar los beneficios ligados al conocimiento.

Lo que aquí se destaca es un caso entre muchos de las complicadas relaciones entre las recomendaciones de los expertos y las decisiones gubernamentales (Nutley, Walter y Davies, 2007). Se vincula asimismo con la distinción entre políticas explícitas e implícitas (Herrera, 1975), frecuentemente visible por ejemplo cuando programas de promoción de CTI nacional son desvirtuados, más allá de intenciones, por disposiciones generales de tipo macroeconómico. Mucho mayor es habitualmente la incidencia en la coyuntura de la macroeconomía que de la CTI; ahora bien, cualquier política de largo plazo es ante todo una cadena de decisiones de corto plazo, en la cual siempre es urgente evitar que se rompan sus eslabones. El Estado es, entre otras cosas, una arena de conflictos entre intereses y actores múltiples; el poder de las "partes interesadas" –*stakeholders*– en CTI es escaso en el subdesarrollo; si no se lo expande, las mejores recomendaciones pueden quedar en letra muerta. Como la política incluye a las políticas públicas, pero no se reduce a ellas (cosa que el inglés

ayuda a recordar por el uso de dos palabras distintas, *politics* y *policies*), las recomendaciones de políticas en CTI resultantes de los análisis de la realidad deben tener en cuenta cómo ampliar su viabilidad política.

Por el sendero esbozado en el párrafo anterior se vuelve a encontrar la cuestión clave de la protección de los aprendizajes. El maestro Hirschman (1958) enseñaba que el enfrentamiento al subdesarrollo consiste ante todo en encontrar recursos descuidados y ponerlos a jugar en pro del desarrollo. En este sentido, son muchos los procesos innnovativos e interactivos de aprendizaje que no son tenidos en cuenta, entre otras cosas por desconocimiento: democratizar incluye también valorar lo que elitistamente no ha sido valorado antes. Detectarlos y sugerir cómo protegerlos debiera ser una de las tareas cardinales de los análisis de las actividades de CTI. En ellos hay experiencias y grupos involucrados cuya incidencia puede potenciarse, incluso en lo que tiene que ver con su peso en las discusiones públicas y las decisiones que van modelando las políticas en CTI.

Tales políticas se definen, por supuesto, ante todo a nivel de los gobiernos. Pero no solo allí, lo que tiene particular importancia cuando los gobiernos no pueden o de hecho no quieren otorgar a las actividades en CTI la jerarquía que sus analistas entienden obviamente merecida o, incluso cuando lo hacen, no logran darle la continuidad imprescindible para que una serie de decisiones se convierta realmente en una política. En cualquier caso, hay ámbitos no gubernamentales en los que la relevancia y la continuidad de las acciones vinculadas con el conocimiento avanzado inciden considerablemente en lo que hace al respecto el país involucrado. En el Norte, ejemplos notorios de ellos son ciertas grandes empresas y también varias universidades. En Latinoamérica, como ya se apuntó, algunas universidades públicas se destacan objetivamente en ese sentido, de modo que el análisis de las políticas en CTI sería incompleto si no las considerara.

Las observaciones formuladas en esta sección de conclusión tienen que ver con una faceta clave de la democratización del conocimiento, como lo es la construcción democrática de estrategias nacionales de largo plazo en CTI. La tarea incluye reconocer intereses y aportes frecuentemente postergados, así como contribuir al intercambio de ideas con elementos de juicio sólidos, orientados a la articulación de intereses diversos. El papel que en ello puede y debe tener la discusión pública democrática lo subraya Amartya Sen al referirse a "la gloria del razonamiento público

abierto, que influencia tanto al conocimiento y a la tecnología como a la política" (2003; nuestra traducción).

Bibliografía

Arocena, R. y Sutz, J. (2020). "The need for new theoretical conceptualizations on National Systems of Innovation, based on the experience of Latin America". *Economics of Innovation and New Technology*, vol. 29, n° 7, pp. 814-829.

Banerjee, A. V. y Duflo, E. (2019). *Good Economics for Hard Times*. New York: PublicAffairs.

Bourdieu, P. (2001). *Science de la science et reflexivité*. Paris: Raisons d'Agir.

Herrera, A. (1975). "Los determinantes sociales de la política científica en América Latina. Política científica explícita y política científica implícita". *Revista Redes*, n° 5.

Hirschman, A. (1958). *The Strategy of Economic Development*. New Haven: Yale University Press.

Nutley, S.; Walter, I. y Davies, H. (2007). *Using evidence: how research can inform public services*. Bristol: The Policy Press.

O'Donnell, G. (2004). "Ciencias sociales en América Latina. Mirando hacia el pasado y atisbando el futuro". *LASA Forum*, vol. 34, n° 1, pp. 8-13.

Sen, A. (4 de octubre de 2003). "Democracy And Its Global Roots". *The New Republic*.

Srinivas, S. y Sutz, J. (2008). "Developing countries and innovation. Searching for a new analytical approach". *Technology in Society*, vol. 30, n° 2, pp. 129-140.

II.
Desafíos para la investigación
¿Ciegos o con perspectiva de género?

Nora Goren

La formulación, el desarrollo y el análisis de los procesos de ciencia, tecnología e innovación (CTI) requieren estar a la altura de las demandas y necesidades de las sociedades en las que se inscriben. Se espera que contemplen de manera dinámica los aspectos económicos, políticos e institucionales y la interdependencia sistémica entre los diversos actores/instituciones que integran dicho sistema. Así, las empresas, las agencias gubernamentales y la sociedad civil van dándole forma a su gestión y evolución (Dutrénit y Natera, 2017).

Las sociedades en las que se inscriben estos procesos de CTI son los espacios en los que van emergiendo cada vez con mayor o con menor intensidad temas en la agenda pública que, en el caso de las demandas instaladas por el movimiento de mujeres y feminista, son una realidad que ha impreso en tales espacios distintos matices y que ya no es posible, o no debería ser posible, ser dejada de lado.

La inclusión del género en la agenda de CTI es disímil según el país al que hagamos referencia; lo que es claro es que queda aún un largo camino por recorrer, sobre todo respecto de cómo y de qué manera se la contempla y se la incluye. ¿Podría ser acaso transversalizar la perspectiva de género o feminista?, ¿es hablar con lenguaje inclusivo?, ¿es hablar de mujeres y varones y disidencias?, ¿es sumar el componente mujeres o cuerpos feminizados a espacios donde antes no estaban?, ¿es incluir bibliografía escrita por mujeres y disidencias?, ¿es construir nuestro objeto de estudio desde una perspectiva que contemple las desigualdades sexogenéricas?, ¿es pensar las respuestas de manera unívoca para toda la población?, ¿es un enfoque epistémico diferente?

Sin pretensión de dar una respuesta unívoca, podríamos decir que son todos y cada uno de esos factores. Lo que sí resulta central es la incorporación de esta mirada/dimensión analítica, desde el momento mismo del *planteamiento del problema*,[1] dado que, más allá de las manifestaciones que concretamente asuma el orden de género –que varían históricamente y de acuerdo con cada sociedad–, estas indefectiblemente están presentes y se expresan en las prácticas, los discursos y los sentidos que les atribuimos a las feminidades y masculinidades, y se hallan por detrás de nuestras definiciones, las hagamos o no explícitas. Y son estas las que se encuentran cruzadas por un sistema de reglas implícitas y explícitas, fuertemente institucionalizadas, vinculadas a cómo organizar la forma de ver y de pensar lo femenino y lo masculino y, por lo tanto, de cómo definir nuestro objeto de estudio. Es allí que el grado de naturalización del que goza este esquema dificulta la puesta en cuestión de desigualdades fácilmente observables si se adopta una mirada de género o feminista, por lo cual es central tenerla presente al momento de definir nuestro objeto de estudios y/o de intervención.

En línea con lo antes señalado, el objetivo de este trabajo es presentar algunas nociones básicas que sirvan para pensar los lineamientos generales de lo que deberíamos tener presente al construir/problematizar nuestro objeto de estudio, sin dejar de tener en cuenta que también puede ser orientador para pensar el proceso en su integralidad, aun cuando cada etapa contiene especificidades diferentes.

Para ello proponemos un recorrido por el concepto de género y sus matices; veremos el lugar que ocupa el lenguaje de género para repensar nuestra forma de mirar e interpelar el mundo y qué implica transversalizar para poder hacer un abordaje integral del problema a abordar, finalizamos con algunos posibles ejemplos y preguntas que pueden ayudar a ver dónde estamos situados/as.

Una aproximación al concepto de género y sus matices

El concepto de género ha tenido un desarrollo histórico que se ha ido complejizando. Así, la distinción entre los conceptos de *sexo* y *género*,

1 Si bien en el planteo del problema se dirimen cuestiones de método, metodología y epistemología, en este artículo haremos una mirada integral, ya que tiene como pretensión dar cuenta de un primer abordaje general.

expresada en el paradigmático "no se nace mujer, se llega a serlo", de Simone de Beauvoir (*El segundo sexo*, 1949), ha permitido alumbrar el carácter social, histórico y contingente de las construcciones de género. El *sistema de sexo-género* (Rubin, 1998) presente en nuestra sociedad nos permite dar cuenta del conjunto de prácticas y sentidos inscriptos en un sistema que dicotomiza y, en consecuencia, define los contornos de lo femenino y de lo masculino, en los cuales las relaciones de género dominantes simultáneamente se producen y se reproducen. Las construcciones de masculinidad y feminidad resultantes se encuentran fuertemente atravesadas por relaciones de poder y autoridad, y así configuran grupos sociales de *mujeres* y de *varones* ubicados según cierto ordenamiento asimétrico fundado en asociaciones entre masculinidad, autoridad y dominio y feminidad, docilidad y abnegación. En palabras de Joan Scott (1996: 34): "El núcleo de la definición reposa sobre una conexión integral entre dos proposiciones: el género es un elemento constitutivo de las relaciones sociales basadas en las diferencias que distinguen los sexos y el género es una forma primaria de relaciones significantes de poder".

Es así como operan diversos dispositivos que asignan valores diferenciales en función del *sexo* de las personas, sexo concebido como mero sustrato material que sirve como base para el *género*, entendido como mutable y social. Al decir de Marta Lamas: "No es lo mismo el sexo biológico que la identidad asignada o adquirida; si en diferentes culturas cambia lo que se considera femenino o masculino, obviamente dicha asignación es una construcción social, una interpretación social de lo biológico" (1986).

Asimismo, no es posible universalizar la categoría mujer. No es lo mismo ser mujer blanca que ser indígena o afrodescendiente; ser joven, adulta o de la tercera edad; residir en zonas urbanas o rurales; pertenecer a un sector social o a otro; vivir en el país de origen o ser migrante; tener o no tener hijos/as.

En esta dirección, los aportes de los feminismos negros y de las mujeres de color, pos y descoloniales dieron lugar a un proceso de robustecimiento de diversas concepciones que buscaban desentrañar el entrecruzamiento de las dominaciones en su complejidad. Es así que la lucha de los colectivos de lesbianas y homosexuales, de bisexuales, travestis, transexuales y personas transgénero, más tarde, se condensaron en una rotunda crítica a los efectos normalizadores y naturalizantes del par sexo/género y proliferaron en la academia a partir de los planteos de pensadoras como la filósofa norteamericana Judith Butler (2002). El punto nodal de la crítica radicó en el esencialismo binario que supone la existencia de

una naturaleza concebible por fuera de las relaciones sociales, encarnada en la distinción macho/hembra y su correlato masculino/femenino, así como en las operaciones epistemológicas concomitantes.

Así, el recorte de género conforma un sistema de referencia, de percepción y organización material y simbólica de la vida social, en el que no es posible encontrar espacios que no se encuentren atravesados por estas concepciones. No se trata únicamente de una cuestión de roles o de funciones, sino que la totalidad de las relaciones sociales está, desde su origen, marcada por el género, y se halla inscripta en lógicas de poder que diagraman posicionamientos jerárquicos establecidos entre los conjuntos sociales de personas divididas según su sexo asignado y categorizaciones sociales vinculadas a la raza y la clase (Goren, Prieto y Figueroa, 2018).

¿Es importante el lenguaje no sexista?

Una vez acordado que varones y mujeres, feminidades y masculinidades, estamos atravesados/as por múltiples desigualdades sintetizadas en lo que denominamos "patriarcado", cabe preguntarnos si podemos hablar, formular un problema desde el denominado lenguaje neutro, o más bien preguntarnos si existe el lenguaje neutro y qué es lo que esa supuesta neutralidad anula y no permite problematizar.

Si lo que no se nombra no existe, si las palabras no solo transmiten el pensamiento, sino que también lo moldean y lo transforman; y si a menudo las normas de género no están explicitadas y se transmiten de manera implícita a través del lenguaje, de las instituciones y de otros símbolos, de la misma forma, un lenguaje específico de género influye en la manera en que se piensan o se dicen las cosas; también en el modo que asumen las relaciones entre mujeres y varones en el mundo y en cómo definir, caracterizar, nuestro objeto de estudio.

Entonces, dar cuenta de que las personas –y en este caso, las feminidades y masculinidades– en nuestras sociedades no somos iguales, por lo que nombrarnos sin invisibilizar la diferencia resulta necesario; hasta me atrevo a decir que no hacerlo debería ser considerado y evaluado de manera negativa, ya que todo el proceso que se continua estará vedado de omisiones.

¿Qué es la transversalización de la perspectiva de género?

Si tomamos como punto inicial la definición adoptada por el Consejo Económico y Social de Naciones Unidas en el año 1997, se entiende por transversalización (ECOSOC, 1997)

> al proceso de examinar las implicaciones para mujeres y varones de cualquier tipo de acción pública planificada, incluyendo legislación, políticas y programas, en cualquier área. Asimismo, es una herramienta para hacer de los intereses y necesidades de varones y mujeres una dimensión integrada en el diseño, implementación, monitoreo y evaluación de políticas y programas en los ámbitos políticos, sociales y económicos.

Como se ha podido apreciar, la transversalización no implica acciones puntuales, sino redefinir todas las actuaciones llevadas a cabo para que estas contribuyan activamente a la igualdad y equidad de género. Con la incorporación de esta estrategia, se adoptaría un enfoque estructural y transformador que conduciría todas las acciones hacia el aporte en esta dirección, sin ser estas acciones específicas, sino integradas a un modo de abordar los temas, problemas, acciones, propuestas. Esto, a su vez, supone contar con las herramientas necesarias para poder alcanzarlo, para el caso de nuestro artículo, tener presentes las desigualdades sexogenéricas como un aspecto constitutivo de lo social y no como un componente que se suma a las propuestas.

Transversalizar supone el ejercicio de un esfuerzo por superar la tradicional separación o segregación de pensar por separado las poblaciones de modo de superar de manera integral las discriminaciones. Y transversalizar la perspectiva de género es el proceso de valorar las implicaciones que tiene para los varones y para las mujeres cualquier acción que se planifique, ya se trate un proyecto de investigación o de una propuesta de políticas o programas, pero lo que es central es pensarla para todas las áreas y en todos los niveles. Con estas especificidades lo que se busca es que la equidad, como principio de justicia, tenga lugar. La igualdad refiere a un principio que establece la igualdad de derechos y oportunidades entre hombres y mujeres tomando como referencia la desigualdad, así como la diferencia de individuos y sus entornos locales. Por su parte, la equidad es un principio de justicia. Esto implica entonces aspirar a la igualdad a partir de la equidad (Ortiz, 2013: 10). Por ejemplo, si tenemos en cuenta la situación diferencial de mujeres y varones en el

mercado de trabajo, estas acciones deberían buscar, por todos los medios, desandar los mecanismos que las perpetúan, para que tanto varones como mujeres sean parte de su abordaje y las mujeres no sean ubicadas como un grupo específico sobre el cual se debe actuar.

Algunos ejemplos

No caben dudas de que la pandemia de covid-19 ha afectado a toda la población en múltiples dimensiones. Una de ellas es el trabajo y, ahora, bien como estudiantes o investigadores, nos solicitan elaborar un diagnóstico que sea de utilidad para pensar posibles respuestas en materia de política pública y que sirva para el diálogo con el sector empleador y el sector sindical. Ahí me enfrento a tener que definir cómo recorto a la población; la primera pregunta esperable sería entonces, ¿puedo hablar de trabajadores en general, cuando sabemos que la participación de las mujeres en el mercado laboral es sustancialmente diferente a la de los varones, ya que no se insertan en los mismos sectores de actividad, no hacen las mismas tareas en el interior de los establecimientos ni son contempladas en las normativas laborales o en los convenios colectivos de trabajo de igual manera (Goren y Trajtemberg, 2018)? Por cierto, podría, pero no estaría dando cuenta de estas diferencias y seguramente las propuestas que puedan surgir de mi evaluación reproducirán las desigualdades que como sociedad nos hemos comprometido modificar. Si en el recorte desde el cual parto no doy cuenta mínimamente de los datos desagregados por sexo, estaré incurriendo en errores de información para quienes tienen la responsabilidad de formular políticas. Asimismo, si no miro otras dimensiones de la vida social, como el reparto de tareas que hacen a la reproducción de la vida cotidiana (léase, cómo se expresa la división sexual del trabajo en las unidades familiares o en los espacios sociales), seguramente no podré dar cuenta de si la carga laboral fue similar o no y qué tipo de carga física o psíquica han debido afrontar unos y otras. Allí deberé enfrentarme a las representaciones hegemónicas por las cuales se supone que las mujeres naturalmente son cuidadoras, y así desconocer todo tipo de carga que ello puede acarrear.

En línea con las representaciones es que también recurrimos a reflexionar de qué manera pensamos el diseño de herramientas industriales y las capacitaciones para su uso. Me debería preguntar acerca de las diferencias sexogenéricas para el diseño de estas herramientas, si son solo

objetos, pero son objetos que luego son utilizados por trabajadores/as. Allí me pregunto: ¿mujeres y varones tienen la misma contextura física, sus manos tienen las mismas dimensiones? ¿Para quiénes están pensadas las herramientas? Asimismo, en las capacitaciones, ¿tenemos en cuenta las historias previas recorridas por unos y otras, para que estas sirvan o tengan en cuenta esa compensación?

¿Cómo saber si estoy considerando el enfoque de género?

En este último punto queremos compartir algunas consideraciones que pueden servirnos a modo de aproximación para saber si efectivamente estamos incorporando la perspectiva de género en el diseño de la investigación y en el análisis de la información disponible. Estas reflexiones no pretender ser exhaustivas, son solo orientativas para posibles preguntas que nos pueden dar una pista de si estamos recorriendo el sentido propuesto. A tales efectos proponemos una serie de preguntas que contemplan dimensiones que, a modo de proxi, pueden ser de utilidad para los fines propuestos.

En una primera etapa, que dé cuenta del *diseño de la propuesta,* se puede observar si se están teniendo en consideración los intereses y necesidades de los grupos afectados por el problema a abordar; de qué manera se están considerando los impactos diferenciales según el grupo específico al que pertenece la población sobre la que se está trabajando o qué impacto tendrá la propuesta a realizar y cómo estos grupos han sido contemplados en la definición de los objetivos planteados, tanto los generales como los específicos.

Luego, cuando nos acercamos al momento del *análisis de la información,* debemos tener en cuenta cómo hemos construido nuestro marco analítico de abordaje, cómo y desde dónde hemos encarado esta perspectiva; si hemos podido identificar las diferencias de oportunidades, de derechos y de acceso de recursos según el género, si hemos tenido en cuenta los aportes diferenciados según el género al tema abordado; o si hemos utilizado indicadores desagregados por sexo.

Otras posibles dimensiones para tener en cuenta: cuál es la distribución de roles en el interior del propio equipo de trabajo y cómo se seleccionan las citas bibliográficas (lugar de la subjetividad en la selección).

A modo de cierre

A lo largo de estas páginas, hemos hecho un recorrido por definiciones básicas que nos permiten acercarnos a conocer de qué se habla cuando hablamos de desigualdades sexogenéricas y cómo considerarlas al momento de definir nuestro objeto de estudio, hasta proponer una serie de preguntas que nos permita ir definiendo nuestros recortes en el proceso de trabajo, desde dimensiones analíticas hasta abordajes epistémicos.

Es claro que la ciencia, tecnología e innovación deben estar a la altura de las circunstancias en todos los planos y darse el tiempo que requiere contemplar cómo es su aporte para reproducir o transformar el mundo, en el que las desigualdades sexogenéricas están en su base constitutiva; es una tarea central que debe estar presente desde el inicio del proceso. Solo de esa manera podremos estar a la altura de los desafíos que la época nos impone.

Bibliografía

Butler, J. (2002 [1993]). *Cuerpos que importan. Sobre los límites materiales y discursivos del "sexo"*. Buenos Aires: Paidós.

De Beauvoir, S. (1949). El segundo sexo, vol. I. *García, Juan (Trad.). Reedición. Buenos Aires: Sudamericana.*

Dutrénit, G. y Natera, J. M. (eds.) (2017), *Proceso de diálogos para la formulación de políticas de CTI en América Latina y España*. Buenos Aires: CLACSO.

ECOSCOC. (1997). Gender Mainstreaming Extract from Report of the Economic and Social Council for 1997 (A/52/3, 18 September 1997). Recuperado de https://www.un.org/womenwatch/daw/csw/GMS.PDF

Goren, N.; Prieto, V. L. y Figueroa, Y. (2018). "Apuntes feministas sobre género y trabajo para pensar la intervención desde el Trabajo Social". *Ts Territorios. Revista de Trabajo Social*, vol. 2, nº 2, pp. 115-128. Disponible en: http://cjys.unpaz.edu.ar/sites/default/files/Ts_2%284%29.pdf.

Goren, N. y Trajtemberg, D. (2018). "Brecha salarial según género. Una mirada desde las instituciones laborales". *Análisis*, vol. 32.

Disponible en: https://library.fes.de/pdf-files/bueros/argentinien/14882.pdf.

Harding, S. (ed.) (1987). "Is There a Feminist Method?". *Feminism and Methodology*. Bloomington, In.: Indiana University Press.

Lamas, M. (1986). "La antropología feminista y la categoría género". *Revista Nueva Antropología*, vol. 8, nº 30, 173-198.

Ortiz, A. (2013). *Modelos Pedagógicos y Teorías del Aprendizaje / Pedagogic Models and Learning Theories*. Bogotá: Ediciones de la U.

Rubin, G. (1998 [1975]). "El tráfico de mujeres: Notas sobre la 'economía política' del sexo". En Navarro, M. y Stimpson, C. R. (comps.), *¿Qué son los estudios de mujeres?* México: Fondo de Cultura Económica.

Scott, J. W. (1996). "El género: una categoría útil para el análisis histórico". En Lamas, M. (comp.), *El género: la construcción cultural de la diferencia sexual*, pp. 34-35. México: PUEG.

Capítulo 1
Análisis del proceso de innovación en América Latina a partir de la combinación metodológica de la modelación basada en agentes y el análisis de redes sociales

Juan F. Franco-Bermúdez, Walter Lugo Ruiz-Castañeda,
María Luisa Villalba-Morales

Introducción

La formulación de políticas públicas de ciencia, tecnología, innovación y sociedad (CTIS) en el contexto de América Latina (AL), considerada una región de economías emergentes (Banco Mundial, 2019), cobra un papel protagónico para el crecimiento económico y para garantizar que los países en desarrollo sean partícipes de las dinámicas globales de índole comercial, científico, tecnológico, ambiental, multicultural, entre otros (Velasco, 2016). Este contexto también está enmarcado por limitantes asociados con la atención de necesidades de un nivel más básico, la mayoría de estas naciones tiene retos importantes en temas de pobreza, alimentación, cobertura de servicios públicos, transporte, salud, educación, orden público, entre otros (de la Cruz Flores, 2017). Si bien estas problemáticas pueden ser apalancadas desde las políticas CTIS, no dan espera para su atención, condicionando las prioridades de la agenda política de las naciones.

Los tomadores de decisiones son los que enfrentan este panorama (Bonvecchi y Scartascini, 2020), ya sean hacedores de política o líderes empresariales, los cuales necesitan apoyarse en marcos que reflejen y

desentrañen esta complejidad, para así soportar sus decisiones de una forma más expedita. Un enfoque metodológico apropiado para ello es la modelación y simulación, lo cual ofrece un abanico de metodologías posibles para comprender sistemas y fenómenos sociales (Epstein, 2008); entre ellas, la modelación basada en agentes (MBA) y el análisis de redes sociales (ARS).

Estas dos metodologías son reconocidas por su capacidad para representar, analizar y proyectar fenómenos sociales complejos (Frantz, 2012), permitiendo descifrar comportamientos no lineales e impredecibles (Aziza *et al*., 2016). Consecuentemente, en este capítulo se explora la combinación de ambas técnicas, las cuales se han reconocido como complementarias y poderosas para abordar problemáticas en el contexto latinoamericano, siendo fundamental para ayudar en la comprensión de las condiciones que afectan el proceso de innovación y poder incidir en una mejor formulación de políticas públicas y estrategias empresariales acordes para estos países. Esto sustentado en la potencialidad del uso de metodologías híbridas (Williams, 2022).

En este capítulo se presenta una descripción general de las metodologías MBA y ARS de manera independiente (incluyéndose la descripción de algunos softwares específicos de cada metodología), para luego converger en cómo realizar una combinación apropiada y contextualizada en AL. Posteriormente, se presenta la discusión en cuanto a los alcances, limitaciones y recomendaciones de cara a la aplicación del método, se exponen algunos casos de utilización de la combinación propuesta y, por último, se plantean condiciones para una posible agenda de trabajo futura.

Descripción de los métodos

Generalidades de la MBA y el ARS

El ARS y la MBA son dos enfoques metodológicos que se han utilizado para analizar diferentes fenómenos que están relacionados con sistemas complejos. Un sistema es complejo cuando sus actores son heterogéneos e interactúan entre sí, luego dichas interacciones producen comportamientos emergentes que son diferentes de los efectos de los elementos individuales y, finalmente, dichos comportamientos pueden persistir en el tiempo o tener procesos de adaptación a situaciones cambiantes en el entorno (Moriello, 2013). Ejemplos de ello se pueden encontrar en estudios

sobre epidemiología, comportamiento social, adopción de tecnologías, comportamientos de mercados, redes sociales, estudios de segregación, entre otros en los que el centro de las investigaciones sean los fenómenos sociales complejos.

En este mismo ámbito se encuentran los sistemas de innovación, los cuales cuentan con las características que les permiten ser considerados como sistemas complejos (Mitchell, 2009) y por ello son propicios para ser estudiados a través de metodologías de MBA y ARS. Cada metodología tiene un alcance y un propósito diferente. Por un lado, la MBA tiene fortalezas para representar la toma de decisiones de los diferentes agentes y replicar los patrones de comportamiento que emergen a partir de la interacción de varios agentes, permitiendo el estudio de la evolución y emergencia del sistema (Jianhua *et al*., 2008); mientras que el ARS permite individualizar los agentes para entender sus oportunidades y limitaciones de acuerdo con los vínculos sociales que mantienen con otros actores (Dahesh *et al*., 2020), y al mismo tiempo analizar las estructuras relacionales complejas presentes en el sistema a diferentes niveles (Freeman, 2004; Luke y Stamatakis, 2012).

Por lo tanto, aquellas investigaciones que tengan como propósito estudiar los patrones emergentes de un sistema basado en las relaciones existentes entre los actores heterogéneos, sus atributos y reglas de decisión que los componen, lo pueden hacer a través de la combinación de la MBA y el ARS, tal como es el caso de Franco-Bermúdez y Ruiz-Castañeda (2019), quienes estudiaron la emergencia de los sistemas de innovación inicialmente con MBA y aumentaron luego el alcance del entendimiento del sistema desde su estructura individual y de red con el uso del ARS, permitiendo identificar elementos claves sobre los agentes (nodos), sus relaciones y el sistema, los cuales no son posibles empleando las metodologías de forma separada.

Aplicación de la MBA y el ARS

Es importante tener en cuenta que cada método se aplica de forma independiente, primero se debe utilizar la MBA para generar los datos que serán utilizados en el ARS; sin embargo, al realizar un análisis conjunto de los resultados arrojados por cada metodología, estos se complementan, pudiendo identificar puntos de apalancamiento en el sistema de innovación, de modo que el análisis sirve de orientación para las decisiones

de estrategia y política pública que tratan de promover los sistemas regionales, sectoriales y nacionales de innovación.

MBA

La MBA tiene por objetivo describir un sistema desde la mirada de los elementos que lo constituyen (Bonabeau, 2002) y estos se consideran entidades autónomas para la toma de decisiones (agentes). Por lo tanto, cada agente realiza evaluaciones individuales de las situaciones de su entorno y toma decisiones basado en un conjunto de reglas.

Con base en ello, su aplicación es apropiada para estudiar la emergencia del comportamiento agregado de un sistema (Rahmandad y Sterman, 2008), que se somete al proceso de modelado y posteriormente se programa, para lo cual, se usa algún software específico (algunos se detallan más adelante). Esto con el fin de realizar experimentos mediante simulaciones orientadas a dar respuestas a preguntas del tipo de ¿qué pasaría si...?

La característica representativa de esta metodología radica en su enfoque, de lo individual a lo general, por lo cual el modelador define las características y las reglas de decisión de cada uno de los agentes que interactuarán entre sí, sin requerir condiciones generales, tales como la de la demanda, el PIB, entre otras.

El proceso de modelado y simulación se puede lograr a través de diversos pasos (Aguilera Ontiveros y Posada Calvo, 2018), sin embargo, si lo que se busca es modelar sistemas complejos adaptables (SCA), estos pasos son complementados con lo propuesto por Holland (2004) (en relación con la identificación de propiedades y mecanismos de los sistemas). Con base en ello, los pasos descriptos a continuación recogen las recomendaciones y buenas prácticas de diversos autores:

- Realizar una lluvia de ideas en la que participe tanto el grupo modelador como los diferentes interesados en el estudio, en la cual se clarifique lo que se desea modelar (cuál es el sistema) y cuál es la pregunta que se está explorando.
- Contrastar las ideas obtenidas en el paso anterior con marcos conceptuales y teóricos que las logre respaldar, de tal manera que se pueda contar con los sustentos de la delimitación y propósito del modelo

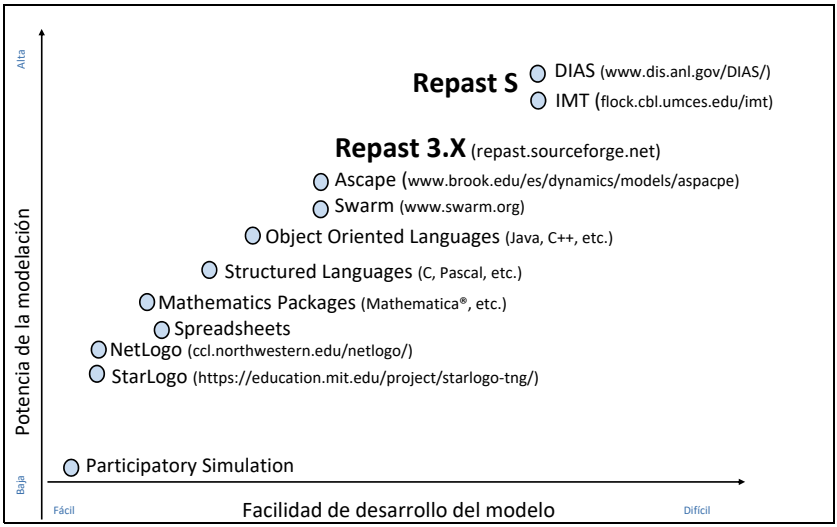
- Identificar las propiedades (agregación, no linealidad, flujos y diversidad) y mecanismos (marbetes o etiquetas, modelos internos y bloques de construcción) del sistema a partir de los dos pasos anteriores. Para mayor claridad en la tipificación de las propiedades y mecanismos, se recomienda revisar a Holland (2004), en cuyo estudio se explica ampliamente el significado de cada uno de estos elementos básicos y cómo su adecuada caracterización permite sintetizar los comportamientos complejos.
- Construir previamente la hipótesis de cómo es el funcionamiento del sistema a partir de la integración de las ideas, teorías, principios y mecanismos en un modelo conceptual, el cual debe permitir el planteamiento de escenarios que buscan cumplir con el objetivo de la investigación. Este paso depende significativamente del software que se vaya a utilizar, puesto que es de acuerdo con su lógica de programación que se describe el sistema. En softwares como NetLogo, es importante definir cada uno de los elementos que componen el sistema, estos son: el ambiente, los agentes (con sus características, variables y reglas de decisión) y la forma en la que interactúan entre ellos. Mientras que, en software como LSD (Laboratory for Simulation Development), por su orientación a objetos, lo importante es identificar las ecuaciones matemáticas que representan las decisiones entre los agentes.
- Realizar un diagrama de flujo que permita describir el proceso de toma de decisiones de cada agente, es decir, que muestre la lógica que tiene el modelo y facilite la programación del código. Una vez que se cuenta con el diagrama de flujo, es posible esquematizar de forma gráfica los cuatro pasos anteriores, para así conservar la visión general del sistema y el propósito del modelo. Este modelo debe ser validado conceptualmente antes de pasar a la simulación.
- Codificar (programar) el modelo. Este paso corresponde a la representación del sistema a través de un conjunto de líneas de texto que equivalen a los procedimientos que debe seguir el programa para representar el modelo y realizar las simulaciones de este.
- Validar el modelo. Esta validación requiere tres momentos, primero se debe verificar el funcionamiento, por el cual, según Sargent (2005), se busca asegurar que la implementación del programa informático del modelo computarizado sea correcta y no contenga

errores de programación. En este paso se realizan simulaciones de prueba en las que se verifique que cada procedimiento realice lo que se había estipulado y que su comportamiento sea consistente en varias simulaciones y condiciones. Luego se realiza la validación tanto conceptual, en la que el modelo debe garantizar una adecuada utilización de conceptos y propuestas teóricas, como operacional, en la que el modelo debe representar el sistema real que se está estudiando, mediante varias técnicas de validación que permiten corroborar la estructura y comportamiento del modelo desde dos frentes, el teórico y el empírico (Ruiz, 2016). Es primordial para que el modelo represente adecuadamente la realidad que los datos empíricos sean ingresados como parámetros al modelo y permitan que se hagan los ajustes necesarios para que así el modelo represente el comportamiento real.

- Realizar las simulaciones de escenarios posibles que permitan estudiar el comportamiento del sistema y contestar la pregunta que se está explorando.
- Ajustar el modelo.

En cuanto a los softwares empleados para la MBA, existe una amplia gama de opciones con diferentes niveles de poder de modelación (representado por las diferentes funciones que permite realizar) y facilidad de uso, tal como se muestra en la figura 1.

Figura 1. Herramientas específicas para la MBA



Fuente: Macal y North, 2006

Uno de los softwares más usado es NetLogo, de acceso gratuito, se caracteriza por ofrecer un entorno de modelado programable y de múltiples agentes, especialmente sistemas complejos que se desarrollan con el tiempo, siendo útil tanto para fenómenos naturales como sociales. Una de sus bondades es que permite explorar las relaciones que se dan en el micronivel y también se puede estudiar e identificar patrones en el macronivel, que resultan de las interacciones entre los individuos (Wilensky, 2020).

Por otro lado, el software LSD no se encuentra en la figura 1. Sin embargo, este software ofrece confianza por ser un sistema de programación automático y permitir la recuperación automática de los datos necesarios para las ecuaciones. Su estructura facilita la programación de modelos relativamente complejos sin requerir conocimientos avanzados. Con este software se puede implementar cualquier tipo de modelo, incluidos los modelos controlados por pares y los modelos basados en estructuras de datos personalizadas, gracias a su arquitectura abierta (Valente, 2008).

ARS

El ARS es una metodología que estudia las interacciones fundamentales de comunicación y cooperación entre los actores partícipes en un sistema (Scott, 1988; Caro, Galindres y Soto, 2013). Por otra parte, Falco, Giandini y Kuz (2016) definen esta metodología como el estudio de la estructura social a partir de las regularidades en el patrón de relaciones establecidas entre entidades sociales específicas como personas, grupos u organizaciones. Estas entidades se conocen como agentes o nodos y tienen asociados determinados atributos, mientras que los enlaces relacionales se denominan vínculos, dichas relaciones determinan unos patrones que dan forma a lo que se conoce como la estructura de la red (Faust y Wasserman, 1994; Borgatti *et al.*, 2009). A partir de todo ello se derivan diversos indicadores que permiten caracterizar el sistema y hacer inferencias sobre este (Granovetter, 1983; Sanz, 2003; Freeman, 2004).

El ARS explica los fenómenos sociales haciendo un estudio en tres vías: 1) el comportamiento individual de cada agente en el micronivel, 2) los patrones estructurales colectivos en el macronivel y 3) las interacciones entre los dos niveles (Sanz, 2003).

Para la construcción de una red se sugiere seguir los siguientes pasos:

- Se eligen cuáles van a ser los nodos que conformaran la red, estos pueden ser individuos, áreas de conocimiento, tareas, actores económicos, etc.
- Luego se elige cuál será el vínculo que conecta a los nodos. Es importante entender que cada vínculo que se elija genera una red diferente. Por ejemplo: en un grupo social el vínculo que genera la confianza (amistad) entre nodos es diferente a la que genera un vínculo de autoridad (jerarquía organizacional).
- Se analizan los resultados del ARS (grafos) mediante sus diferentes indicadores y, si hay necesidad de una mayor comprensión del fenómeno, se procede a repetir los pasos anteriores para construir una nueva red desde otra perspectiva.

A continuación, se presentan los indicadores que comúnmente se utilizan en los estudios de ARS (Freeman, 1978; Faust y Wasserman, 1994; Scott, 2000).

Indicadores de nivel individual

Cuando se analizan los nodos de manera independiente, se analizan medidas de centralidad (Freeman, 1978), las cuales indican qué tan bien (o mal) está posicionado y conectado el nodo en el interior de la red.

- Centralidad de grado: es el número de nodos para los cuales un actor es adyacente, es decir, el número de conexiones directas que tiene un nodo. Refleja la exposición e importancia que tiene un nodo dentro de la red.
- Centralidad de cercanía: indica la capacidad de un nodo de llegar a todos los otros nodos de la red en pocos pasos; este concepto se asocia con la independencia del agente para participar dentro de la red. Un nodo con alto grado de cercanía puede alcanzar con mayor facilidad los recursos que fluyen a través de la red (en términos de información y/o conocimiento, por ejemplo), mientras que un nodo con bajo grado de cercanía depende de otros nodos para poder alcanzar dichos recursos.
- Centralidad de intermediación: es el número de caminos más cortos que pasan por un nodo y posibilitan un vínculo indirecto entre pares de nodos no adyacentes, es decir, cómo un determinado nodo facilita la propagación de información y/o conocimiento sirviendo de enlace a través de la red.
- Centralidad del vector propio o eigenvector: indica cuando un nodo con alta centralidad de grado se relaciona a su vez con otros agentes cuya centralidad de grado también es elevada. Indica una medida de popularidad del agente, que además de tener múltiples vínculos, se relaciona con otros agentes que también están "bien conectados" en el interior de la red.
- Centralidad de poder: también conocida como centralidad de Bonacich (Bonacich, 1987). Es una medida que parte de la idea de la centralidad del vector propio, pero además de indicar si los nodos adyacentes a un actor son populares, refleja si dicha relación es positiva o negativa en términos de poder para el nodo en cuestión. Es decir, si tener vecinos poderosos aumenta o reduce el estatus de poder en el interior de la red.

Indicadores de nivel completo

Cuando se analizan las estructuras relacionales en el macronivel también se analizan medidas de centralidad, pero principalmente se estudian medidas de cohesión, las cuales indican qué tan bien (o mal) está conectada la red, también se conoce como conectividad.

- Grado medio: es el promedio de la centralidad de grado de todos los agentes, entendido como el número de vecinos directos que en promedio tiene un actor perteneciente a la red. Un alto grado medio refleja alta cohesión.
- Índice de centralización: es una medida de centralidad que indica qué tan dependiente es la red de unos pocos nodos que poseen la mayoría de las conexiones y, por tanto, se convierten en nodos centrales. Un alto índice de centralización refleja una red frágil ya que la desaparición de uno de los agentes centrales puede determinar la desaparición de la propia red o gran parte de ella.
- Densidad: es una medida de cohesión que relaciona el número de vínculos presentes en la red con el número de vínculos posibles. A mayor densidad, mayor cohesión; si todos los pares de nodos tienen un vínculo directo, la densidad toma el valor de 1.
- Alcanzabilidad: es una medida de cohesión que relaciona el número total de pares de nodos alcanzables respecto al total de vínculos posibles. Dos nodos son alcanzables entre sí, aunque no sean adyacentes, si existe un camino entre ellos que los conecte. Si cualquier nodo puede alcanzar cualquier otro nodo de red, la alcanzabilidad toma el valor de 1.
- Transitividad o coeficiente de clusterización: es una medida de cohesión que relaciona el número de triángulos (relación recíproca entre tres nodos) presentes en la red, respecto al número de tríadas (grupo de tres nodos alcanzables entre sí). Si una tríada forma a su vez un triángulo, se reconoce como un clúster. Una alta transitividad no necesariamente refleja alta cohesión; indica que la red se hace menos dependiente de los nodos centrales, pero puede tener lugar en la existencia de múltiples comunidades o subgrupos con alta cohesión en el interior de estos pero baja cohesión con el resto de la red.

- Modularidad: es una medida de aglomeración, es decir, que indica si la red se puede dividir en particiones o comunidades cuyos nodos comparten características entre sí y están bien conectados, pero que difieren de las características de los nodos pertenecientes a otros subgrupos. Si se analiza cada comunidad de manera independiente como una nueva red, se observarían altas densidades y altos coeficientes de clusterización. Este concepto viene de la idea de agrupamiento jerárquico en análisis de datos.

En cuanto a los softwares específicos, a continuación, se presentan generalidades de los paquetes más usados para análisis de redes sociales.

Inicialmente, Gephi es un software de manipulación intuitiva con una interfaz robusta, es particularmente poderoso para visualización de redes permitiendo incluso vistas en 3D y la construcción de grafos dinámicos, es consistente con redes de alrededor de veinte mil nodos, sin embargo, es limitado en cuanto a componentes de analítica y estadística (Bastian, Heymann y Jacomy, 2009). Pajek, aunque no tiene tantos atributos gráficos, es eficiente para la generación de redes grandes (mayores a cien mil nodos), permitiendo crear redes aleatorias según parámetros ingresados por el usuario (Batagelj y Mrvar, 1998). Luego, VOSViewer fue desarrollado particularmente para análisis bibliométricos, siendo compatible con datos exportados directamente de fuentes como Scopus y Web of Science, incluye la funcionalidad de minería de texto, es bueno para la visualización de la red, pero no para propósitos analíticos, automatizaciones o el análisis de otro tipo de redes (van Eck y Waltman, 2010).

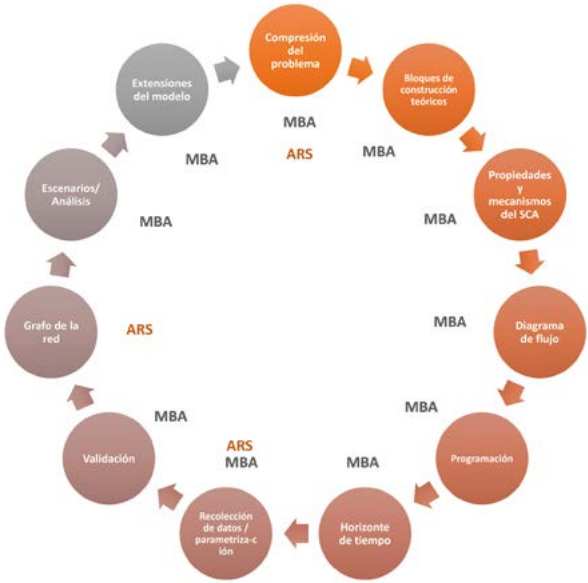
UCINET fue uno de los primeros en aparecer y es uno de los más populares, aunque el desempeño se ve afectado con redes entre los cinco y diez mil nodos, permite trabajar con hasta treinta y dos mil nodos, sin embargo, permite exportar datos a Pajek. Si bien no es tan intuitivo y no tiene atributos de visualización tan potentes como los de Gephi, está muy bien documentado y tiene características favorables para el procesamiento de datos (Borgatti *et al.*, 2009).

Finalmente, el paquete R/igraph es una de las mejores herramientas, esta librería permite graficar las redes, pero la mayor ventaja de esta integración es que pueden aprovechar las fortalezas analíticas y estadísticas de R para el procesamiento de datos y las automatizaciones; la principal limitación es que requiere conocimientos en el lenguaje R siendo más accesible para programadores y estadísticos (Luke, 2015: 4-6).

Procedimiento sugerido para la combinación de las metodologías

Como se presentó en los apartados anteriores, existen pasos para la utilización de cada una de las metodologías, la combinación de ellas debe estar compuesta por estos de forma integrada (ver figura 2), los cuales se describen a continuación, adicionando una mirada desde los problemas de CTIS en el contexto de AL.

Figura 2. Pasos para la aplicación integrada de MBA y ARS



Fuente: elaboración propia

1. Se necesita una comprensión profunda del problema de CTIS que se quiere estudiar, para ello es necesario tener claridad en la pregunta que se quiere responder mediante la MBA. Luego es necesario responder cómo la experimentación mediante la simulación de diferentes escenarios puede aportar a la comprensión del fenómeno. Y entender por qué a través del ARS, que permite obtener las propiedades relacionales, estructurales, nodales y de la propia CTIS involucrada, se obtendrán resultados diferentes a la utilización de otras metodologías. Para continuar con esta

conceptualización del sistema se requiere hacer explícitas las ideas que se quieren examinar con el modelo; estas ideas se pueden obtener a través de técnicas de ideación como la tormenta de ideas, el método Delphi, el análisis morfológico, las analogías, los cinco por qué, entre otras; esto facilita la construcción posterior del modelo y, especialmente, elegir qué detalles del sistema son esenciales y cuáles pueden ser obviados. Entendiendo que un modelo siempre será una representación limitada de la realidad y querer introducir todos los componentes puede ser contraproducente. Sin embargo, esta elección debe hacerse pensando en cómo el modelo facilitará la comprensión del problema de CTIS identificado.

2. Seguidamente, las ideas que se utilizaron en el apartado anterior deben ser soportadas con bloques de construcción teóricos, considerados como uno de los mecanismos que componen los sistemas complejos adaptables (SCA); permitiendo, mediante la disgregación de los bloques de construcción del problema de CTIS, la extracción de las reglas de decisión necesarias para la MBA y el vínculo o vínculos que serán analizados mediante el ARS.

3. Para continuar afinando los elementos que son necesarios para la construcción del modelo MBA (el cual nos permite abordar la complejidad, fruto de la adaptación de los agentes, y la evolución de los vínculos que se analizarán a través del ARS), es importante identificar las propiedades y mecanismos del problema de CTIS mediante su abordaje como un SCA. A continuación, se presentan las cuatro propiedades y los tres mecanismos propuestos por Holland (2004) que deben ser identificados en el SCA:

- Agregación (propiedad): se debe identificar cuál es el comportamiento emergente, entendido como la característica básica de todos los SCA, que trasciende la conducta individual de los agentes componentes, en el caso de la CTIS en AL esto puede ser gasto en I+D, nivel inventivo, comunidades incluidas en dinámicas innovadoras, crecimiento del PIB, innovaciones inclusivas, etc.
- Etiquetado (mecanismo): facilita la interacción selectiva de los agentes mediante su identificación, permitiendo que en el SCA se presente la discriminación, la especialización y la cooperación, fundamentales para que después se presenten comportamientos

emergentes. En AL, por ejemplo, esto puede ser percibido en la conformación de coaliciones políticas, que conllevan a la formulación de ciertas políticas de CTIS.

- No linealidad (propiedad): dada la heterogeneidad de agentes de un SCA y las diferentes reglas de decisión de cada uno, es poco probable identificar el agregado de sus interacciones, provocando que el comportamiento resultante sea más complicado y difícil de prever mediante simples sumatorias, promedios o ponderaciones comunes en las funciones lineales. Esto es muy común cuando se implementan instrumentos de política de CTIS, en las que el saber su efecto *a priori* es muy complicado. Esta es una de las bondades de este tipo de metodologías, precisamente por su propiedad de no linealidad.
- Flujos (propiedad): es fundamental encontrar la relación entre nodo, enlace y el recurso que fluye a través del vínculo entre ellos. En el caso de la CTIS los flujos más normales son de información, conocimiento, dinero, paquetes tecnológicos, innovaciones, innovaciones inclusivas, etc.
- Diversidad (propiedad): la cual se representa en la heterogeneidad de los diferentes agentes (nodos) del SCA y la evolución de estos mediante la adaptación. Esta propiedad permite representar la gran variedad de actores e intereses presentes en los problemas de CTIS de AL.
- Modelos internos (mecanismo): estos permiten identificar qué compone al agente (nodo) y el medio ambiente (conformado por otros agentes), y cómo estos van coevolucionando a partir del aprendizaje fruto de su experiencia. Una forma común de representar los agentes en problemas de CTIS es a través de sus capacidades y el ambiente; estas pueden ser políticas de CTI que afectan su mantenimiento.
- Bloques de construcción (mecanismo): como se vio en el ítem anterior, son los soportes teóricos del SCA. En el caso de los problemas de CTIS algunos de los bloques de construcción usualmente utilizado son: sistemas de innovación (nacionales, regionales, sectoriales, tecnológicos, localizados, inclusivos), la mirada de recursos y capacidades, modos de política de CTI, economía evolutiva, entre otros.

4. Al tener identificados todos los elementos anteriores se procede con la construcción del diagrama de flujo, en el cual se expresa el orden en que los agentes aplican sus reglas de decisión.

5. Teniendo el diagrama de flujo, se continúa con la programación, en la que se lleva a procedimientos codificados lo representado anteriormente de forma gráfica, con estos se construye el modelo computacional, el cual permitirá realizar experimentos a través de la simulación de diferentes escenarios.

6. De acuerdo con el comportamiento resultante de la evolución de los diferentes actores que componen el sistema de CTIS en el tiempo, se debe establecer el horizonte a simular, esto es clave porque se sabe que los SCA son específicos y no se puede generalizar la velocidad con que se produce la adaptación.

7. Para poder utilizar la MBA, es necesario parametrizar los diferentes elementos que la componen, por lo que se requiere un trabajo empírico que permita la recolección de la data necesaria, para luego comparar el comportamiento de los resultados arrojados por la MBA con la realidad. Es en este paso a partir del cual se concentran las mayores limitaciones de la aplicación de MBA y ARS en el contexto latinoamericano. En el tercer apartado se continúa con esta explicación.

8. Se procede con la validación de la MBA, la cual, como ya se mencionó anteriormente, consta de tres partes: verificación del modelo computacional, validación del modelo conceptual y validación operacional.

- La verificación se realiza sobre los procedimientos codificados para asegurar que el programa informático es consistente.
- La validación conceptual permite establecer que las teorías e hipótesis se han incorporado de forma correcta en el modelo, y la estructura, lógica, relaciones matemáticas y causales se incorporaron en el problema de CTIS de una forma razonable de acuerdo con el objetivo de la construcción del modelo.

- La validación operacional se da cuando, al cargar el modelo con los datos empíricos, este refleja un comportamiento correspondiente al del sistema real.

9. Con el modelo validado y los datos empíricos cargados, se puede realizar el ARS en cualquier momento de la simulación del modelo MBA. Esta decisión temporal es fundamental en el ejercicio de combinar las dos metodologías y se puede decir que es clave para obtener hallazgos novedosos que escapan a otros abordajes metodológicos. Primero, el modelador debe elegir el tiempo en que se realizará la simulación del modelo MBA, este período de tiempo depende del comportamiento del problema, el comportamiento del sistema, la incertidumbre asociada al fenómeno estudiado, etc. Entendiéndose que en este aspecto interviene la habilidad del investigador. Segundo, otra disyuntiva es la elección del momento o momentos en que se debe realizar el ARS, para ello, es importante observar los patrones de comportamiento de las diferentes simulaciones efectuadas, para identificar turbulencias, umbrales, equilibrios, etc. En estos no se tiene claridad en el porqué de ese comportamiento emergente del sistema o punto de inflexión y el ARS puede brindar luces de lo que está pasando a nivel nodal o en la estructura de la red que ocasiona ese comportamiento que se vuelve "invisible" para la MBA, robusteciendo el nivel de análisis y permitiendo hallazgos que favorezcan una mejor comprensión de fenómenos complejos como es la CTIS.

10. Ahora viene el diseño y experimentación con escenarios, los cuales deben permitir contestar la pregunta de investigación original. Si la simple simulación del caso empírico, como el realizado en la etapa anterior, permitió una mejor comprensión del problema de CTIS, la posibilidad de hacerse preguntas del tipo de ¿qué pasaría si...? es fundamental para el diseño de políticas de CTIS, y luego poder analizar momentos de la simulación a partir de la cual se pretende entender el rol de uno o varios nodos específicos, además del efecto de sus relaciones mediante el ARS, esto da un abanico de posibilidades de análisis que puede superar limitantes identificados con otras metodologías que no permiten la experimentación con sistemas sociales complejos.

11. Proponer extensiones del modelo, a partir del cual se quieran involucrar elementos que no fueron tenidos en cuenta en una primera

instancia, permitiendo la continuación en la construcción de hipótesis y conocimiento a partir de un modelo base validado y con unos bloques de construcción sólidos.

Discusión de la información/datos requeridos para la aplicación del método en CTI

Consideraciones de validación/tipos de fuentes de información disponible

Una de las grandes críticas que se les hace, principalmente, a los modelos de MBA es la validación, porque no siempre hay datos empíricos longitudinales que permitan comparar las simulaciones con la realidad, esto se puede convertir en un limitante clave, especialmente en los trabajos que se realizan en AL, puesto que, generalmente, no se posee una base confiable ni mucho menos a través del tiempo de data.

Sin embargo, hay algunas técnicas de validación que pueden ser útiles para superar estos limitantes, que en el contexto de AL son muy comunes. Sargent (2005) provee una caja de herramientas de técnicas que se pueden usar cuando no se poseen los suficientes datos empíricos para apoyar la validación, tanto conceptual como operacional.[1] Sargent también recomienda que se deben utilizar al menos dos técnicas en cada validación (conceptual y operacional) para darle mayor validez al modelo construido. Las siguientes alternativas son recomendadas cuando se tienen esas dificultades:

- El método histórico del racionalismo busca comprobar que los supuestos de un modelo son ciertos; esto se hace a partir de premisas que se desprenden de deducciones lógicas, basadas en la teoría, para desarrollar el modelo válido. Puntualmente, este método implica hacer explícitas esas ideas que fueron contrastadas con la teoría y evidenciar la forma en que fueron operacionalizadas a

1 Sargent presenta quince técnicas de validación (animación, comparación con otros modelos, pruebas degeneradas, validez del evento, prueba de condiciones extremas, validez de cara, validación de datos históricos, métodos históricos (pueden ser del racionalismo, empirismo o economía positiva), validez interna, validación de varias etapas, gráficos operativos, variabilidad de parámetros o análisis de sensibilidad, validación predictiva, trazas y prueba de Turing); sin embargo, se introducirán las técnicas que se consideran más aplicables al contexto de la CTIS para AL, dadas las particularidades ya mencionadas.

través de los supuestos y las reglas de decisión de los agentes del modelo. Esta técnica se utiliza especialmente para la validación conceptual, buscando garantizar que en el modelo conceptual estén correctamente utilizados los bloques de construcción elegidos.

- La aproximación histórica amigable, centrada en la utilización de estudios de casos históricos específicos de una industria, pudiéndose hacer inferencias lógicas para conjeturar los parámetros, interacciones y reglas de decisión de los agentes del modelo; pudiéndose también validar el modelo cuando este puede generar múltiples hechos estilizados observados en una industria, orientándose específicamente al comportamiento, reglas de decisión y las interacciones de los agentes y el entorno en el que operan. Esta técnica se basa principalmente en relatos, siendo lo ideal tener varios relatos o casos. Es utilizada principalmente para realizar validaciones operacionales, en las que se identifican y extraen los relatos que han arrojado estudios previos del sistema real que se quiere representar con el modelo; luego, se comparan los resultados arrojados en la simulación del micromundo, que ha sido alimentado con los parámetros obtenidos mediante el trabajo empírico, con estos relatos identificados previamente. Finalmente, si las simulaciones del modelo parametrizado, con los datos obtenidos del sistema real, logran representar de una forma plausible los relatos a los que se llegó a través de los trabajos previos, se puede decir que el modelo operacionalmente es válido.
- La variabilidad de parámetros o análisis de sensibilidad. Este método consiste en cambiar los valores de entrada y los parámetros internos de un modelo para determinar el efecto sobre el comportamiento o la salida del modelo; validándose cuando las mismas relaciones observadas en el modelo son las que ocurren en el sistema real. Es importante tener en cuenta que esto puede ser usado tanto cualitativa como cuantitativamente, dando un margen de acción muy importante cuando se tienen relatos, pero no datos exactos. En la práctica, los sistemas reales han experimentado cambios en el tiempo fruto de decisiones que se tomaron en diferentes momentos; por ello, para la validación del modelo operacional, es importante obtener relatos y/o datos del sistema real antes y después de ese cambio dado, por ejemplo, una nueva política de CTI por la cual se pasa de una orientación hacia la I+D

a una enfocada en el relacionamiento, para luego parametrizar en el modelo computacional estas diferentes condiciones, esperando que se vean reflejados los comportamientos anteriormente identificados y así poder decir que el modelo es válido.

- La validez de cara (*face validity*) es otra valiosa alternativa para realizar la validación cuando hay escases de datos, en la que se indaga a personas que conocen el sistema y/o su comportamiento si los resultados arrojados por el modelo son razonables al compararlos con lo que el experto esperaría. Esta técnica se puede emplear de varias maneras: presentación a cada experto de los resultados arrojados por el modelo de forma individual para después comparar sus respuestas o la realización de un *focus group* con los expertos en el sistema en estudio, cualquiera de estas dos maneras es posible y en muchas ocasiones se argumenta su utilización a partir de la facilidad de acceso que se tenga a los expertos; adicionalmente, se puede o no, decirles previamente a los expertos que los resultados que serán presentados son fruto de una simulación computacional, el ocultar esta información puede generar apreciaciones muy diferentes por parte de los expertos por el mismo nivel de subjetividad que está presente en esta técnica. Esta técnica puede ser utilizada tanto en la validación conceptual como en la operacional, en el caso de ser utilizada en la validación conceptual, lo que se les presenta a los expertos son los supuestos y reglas de decisión utilizados en la construcción del modelo conceptual.

Es importante tener en cuenta la recomendación de Sargent (2005): aplicar al menos dos técnicas de validación de la caja de herramientas, para la validación conceptual y la operacional (se pueden mezclar como se desee el modelador, teniendo en cuenta el acceso a los datos).

Alcances y limitaciones

La MBA, además de contar con las ventajas generales de la simulación, también puede ofrecer información adicional de un sistema que otros métodos de simulación no ofrecen, como el de dinámica de sistemas, debido a que es capaz de capturar explícitamente la estocasticidad inherente del sistema (Macal, 2010), permitiendo la autonomía, razonamiento,

comunicación y coordinación de los agentes (Dongsheng y Yongan, 2008), y de esta interacción en el micronivel puede emerger el macronivel o agregado (Garcia y Jager, 2011). También permite representar la emergencia y dinámica del sistema, los bucles de realimentación, el carácter heterogéneo de los actores, su evolución y su adaptación (García-Vázquez y Sancho Caparrini, 2016).

Por su parte, el ARS se diferencia de los métodos convencionales, como la estadística descriptiva, porque involucra el carácter relacional de los actores de un sistema identificando agentes de mayor importancia e impacto, formas de colaboración, dinámicas de transferencia tecnológica y difusión de conocimiento, entidades excluidas, entre otros patrones que no agrupa ninguna otra metodología. Adicionalmente, permite dos visuales, una estructural relacionada con la topografía de la red y otra particular para entender las oportunidades y limitaciones de cada nodo en específico. También es una metodología gráfica que permite múltiples opciones de visualización que facilitan la construcción de inferencias y hallazgos; no tiene limitaciones en cuanto al número de agentes y vínculos que se pretenda analizar, también permite observar la evolución de la red en el tiempo, así como patrones de comportamiento (Catebiel, Castro y Hernández, 2006; Aguilar *et al*., 2016).

A través de la combinación metodológica propuesta se superan algunas barreras y se obtiene un efecto ampliado en el entendimiento de cualquier sistema y fenómeno social. Ambos métodos economizan tiempo y recursos; se abre la puerta para realizar múltiples experimentos controlados para entender el efecto de introducir cambios al sistema (Canessa y Quezada, 2010), con la visual del impacto en la estructura relacional y de cooperación del sistema. Se resuelve la principal dificultad del ARS que es la consecución de datos relacionales y el no mapeo de agentes que pueden hacer parte de la red pero son difíciles de identificar (por ejemplo, actores excluidos), un proceso que normalmente es costoso, complejo y de larga duración (Miceli, 2008). Se tiene una visión más amplia que no puede obtenerse al aplicar las metodologías por separado, comprende la emergencia, evolución y comportamientos del sistema dados por la MBA, así como el análisis global e individual proporcionado por los indicadores del ARS, esto da lugar a una asertiva toma de decisiones (Franco-Bermúdez y Ruiz-Castañeda, 2019). AL se convierte en un laboratorio excepcional para realizar este tipo de análisis, puesto que existe un gran número de problemáticas y fenómenos de carácter social que necesitan ser apalancados a través de políticas y estrategias de CTIS (Gaudin y Pérez, 2014).

Finalmente, algunas limitaciones que es pertinente considerar son: 1) los modelos de MBA siempre requieren de validación para garantizar que representan el sistema que se pretende estudiar, una construcción incompleta de los modelos puede dificultar la replicabilidad de los resultados en el sistema real (Canessa y Quezada, 2010), transitoriamente esto puede impactar los resultados del ARS. 2) Particularmente en AL, las metodologías (MBA y ARS) tienen un alto grado de novedad y su implementación es lenta (Gonzáles-Aguilar y Pinto, 2014). De igual forma, la cantidad de cursos que se dictan en universidades es mínima y se cuenta con bajo número de expertos en la aplicación de los métodos. Esto se evidencia en que la mayoría de los estudios realizados son contratados con investigadores de Europa. Para el ARS es crucial una correcta selección de indicadores que realmente respondan al objetivo de la investigación, también es importante entender que dado el carácter social de los sistemas que se estudian, las intervenciones que se realizan tienen efectos directos sobre las personas, allí cobra importancia la ética y responsabilidad social (Miceli, 2008).

Casos de aplicación en América Latina

Los tres casos mostrados a continuación corresponden a trabajos que, con diferentes alcances y propósitos, hacen uso de la combinación de la MBA y el ARS para el análisis de problemáticas tipo, presentes en el contexto de AL: fenómeno de la innovación, atención médica e ingreso al mercado laboral. Cada trabajo se realiza inmerso en limitaciones, con aportes al entendimiento del problema estudiado y abre las puertas a la necesidad de dar continuidad al uso de estas metodologías para el abordaje de problemas complejos.

Caso 1: sistemas de innovación – rol del intermediario

En el trabajo de Franco-Bermúdez y Ruiz-Castañeda (2019), enmarcado en el estudio de sistemas de innovación, se plantea la combinación de estas dos propuestas metodológicas, encontrando que se refuerza la comprensión del fenómeno en estudio. En primer lugar, permite vislumbrar la emergencia del fenómeno, gracias a la MBA, para luego, a través del ARS, profundizar en el comportamiento individual de cada agente y cómo sus características en ciertos períodos de tiempo le permiten

obtener determinados desempeños en los estados futuros. Esto permitió evidenciar la dependencia de la trayectoria y la importancia que tiene la historia en el proceso de innovación; adicionalmente, el ARS permite estudiar la topología de la red, con el fin de medir la eficiencia y su estabilidad, así como identificar los efectos que tienen esas condiciones (como la presencia de comunidades y el rol de nodos intermediarios, por citar algunos ejemplos) en el comportamiento futuro.

Por consiguiente, la combinación de las dos metodologías, especialmente cuando se abordan fenómenos tan complejos, como los efectos de las políticas en CTI en el desempeño de los sistemas de innovación, logran una poderosa herramienta de análisis que logra desentrañar puntos de apalancamiento que son invisibles con otros abordajes metodológicos.

Del ARS es destacable la diversidad y amplitud en la que es posible aplicar el método y la amplia gama de estudios que se han adelantado en múltiples campos como la economía, sociología, ecología, biología, desarrollo sostenible, salud, comunicación y educación (Peyró, 2015).

Caso 2: modelo de diagnóstico in situ de una enfermedad

En el trabajo de Quezada y Canessa (2010) se realizó un modelo de MBA que permite representar el sistema de atención médica en determinadas ubicaciones geográficas donde se encuentran los pacientes, los cuales son diagnosticados y se deben tomar dos posibles decisiones: trasladar al paciente a una instalación médica o tratar a la persona en el lugar. El objetivo era crear una estrategia efectiva para detectar la totalidad de enfermos teniendo un personal reducido y enfermos agrupados en localidades de diferente densidad poblacional, problemática que persiste en AL debido a la distribución y exclusión de algunas comunidades. De este trabajo se concluyó sobre la utilidad de la MBA en la representación y análisis de fenómenos sociales, así como su gran aporte al diseño de experimentos; también concluyeron que el flujo de información entre paramédicos es indispensable y tiene gran influencia en el cumplimiento del objetivo, mientras que la densidad poblacional tiene un efecto reducido sobre este; se entendió el carácter relacional de los agentes del sistema y su importancia.

Paralelamente con el ARS se ha demostrado que en los fenómenos sociales de carácter sanitario es necesario desarrollar una cultura de seguridad basada en una comunicación abierta, incluyendo profesionales, líderes de servicios, las instituciones de salud y los ciudadanos, y que la

efectividad y eficiencia de dicha comunicación depende netamente de la estructura de la red social (Agra-Varela *et al*., 2013). Se evidencia la complementariedad de ambos métodos.

Caso 3: acceso al mercado laboral

El acceso al mercado laboral es un fenómeno social con índices alarmantes en AL. Este fenómeno se caracteriza por una alta dependencia de las redes sociales en las que se encuentren los agentes que buscan una oportunidad laboral, los agentes que la pueden ofrecer y los agentes que pueden conectar al oferente con el demandante. Por lo tanto, García-Valdecasas (2014) usó MBA y ARS para determinar la influencia del tipo de red y las condiciones de desigualdad (representado por el índice de Gini) en la consecución de un puesto de trabajo.

La descripción técnica y las conclusiones de este trabajo están enfocadas principalmente en las características de las redes (densidad, índice de globalización, configuración y la distribución de los vínculos, velocidad de difusión de la información, entre otros), lo que permite identificar una mayor concentración del uso de ARS para la investigación, surgiendo el interrogante de ¿cuál fue el aporte de la MBA en esta investigación? Ante ello, y de acuerdo con la información del documento, la MBA fue usada para suplir la falta de datos, debido a la "dificultad o imposibilidad de conseguir datos empíricos suficientes" (García-Valdecasas, 2014).

De lo anterior, se resalta la complementariedad entre las metodologías, puesto que las simulaciones que se realicen con la MBA pueden generar insumos para aplicar el ARS y así contrarrestar una de las dificultades para aplicar este tipo de metodología (los datos). Sin embargo, es necesario que en los trabajos se expliciten la validación de los modelos empleados, para dar rigurosidad a la obtención de los resultados.

Recomendaciones para futuras aplicaciones

Como lo reconoce Holland (2004), la adaptación de los agentes a los estímulos de su medio ambiente (el cual está conformado por otros agentes) es la que origina la complejidad. Lo anterior obstaculiza los intentos por comprender los fenómenos sociales, como por ejemplo el proceso de innovación. El entorno de AL, en el que difícilmente se formulan políticas

de largo plazo, genera un entorno más inestable, que requiere una adaptación más rápida y por ende aumenta la complejidad.

Es acá que metodologías como la MBA, el ARS y su combinación se presentan como alternativas que permiten comprender mejor esta complejidad, mediante experimentos que no se pueden realizar en la realidad; facilitando de esta manera el entendimiento de las diferentes variables que afectan los sistemas complejos, y así, apoyar la formulación de políticas de CTI con una mayor probabilidad de obtener los efectos esperados y con una mejor utilización de recursos escasos, para finalmente utilizar instrumentos de política que se comporten como puntos de apalancamiento para el desempeño deseado del sistema.

En cuanto al estudio de sistemas de innovación u otros sistemas sociales, los cuales son afectados por las políticas de CTIS, se debe tener presente que las redes se caracterizan por las normas de reciprocidad y confianza que no suelen surgir espontáneamente, requiriendo un esfuerzo deliberado para crear conexiones entre los actores, en el contexto latinoamericano esto tiene sus particularidades que no se deben ignorar y que, incluso, varían en niveles más pequeños como el regional y el municipal (García-Guadilla, 2013). También hay que entender que los nodos en la red tienen un enfoque centrado en su aporte comunitario, lográndose esto a través no solo de los conocimientos del nodo, sino del trabajo comunitario que se fortalece gracias a la participación en la formación de la red en la que intervino dicho nodo, tal participación está fuertemente condicionada por el contexto de AL. Adicionalmente, es difícil medir los efectos comunitarios, generalmente, por falta de una manera de evaluarlos, puesto que, por lo general, estos se deben suponer y se convierten en un acto de fe (Ghoshal y Nahapiet, 1998; Morales y Sifontes, 2014), siendo esto último el problema que se puede resolver con la MBA, para posteriormente aplicar el ARS, y así complementar los hallazgos obtenidos (Franco-Bermúdez y Ruiz-Castañeda, 2019).

El presente capítulo aporta un procedimiento para combinar las metodologías MBA y ARS, y su potencial de análisis, especialmente cuando se quiere comprender el posible efecto de las políticas de CTI y las relaciones entre diferentes agentes en el desempeño de los sistemas de innovación en el contexto latinoamericano. Convirtiéndose así en una poderosa herramienta mixta para investigadores y hacedores de política de CTI que se enfrentan a estos fenómenos complejos que requieren metodologías diseñadas para desenmarañar este tipo de sistemas.

Bibliografía

Agra Varela, Y.; González Pérez, M.; Marqués Sánchez, P.; Pinto Carral, A.; Quiroga Sánchez, E. y Vega Núñez, J. (2013). "El análisis de las redes sociales: un método para la mejora de la seguridad en las organizaciones sanitarias". *Revista Española de Salud Pública*, vol. 87, nº 3, pp. 209-219.

Aguilar, J.; Aguilar, N.; García, E.; Martínez, E.; Muñoz, M. y Santoyo, H. (2016). "Análisis de redes sociales para catalizar la innovación agrícola: de los vínculos directos a la integración y radialidad". *Estudios Gerenciales*, nº 32, pp. 197-207. DOI: https://doi.org/10.1016/j.estger.2016.06.006.

Aguilera Ontiveros, A. y Posada Calvo, M. (2018). *Introducción al modelado basado en agentes. Una aproximación desde NetLogo.* San Luis Potosí: El Colegio de San Luis.

Aziza, R.; Borgi, A.; Guinhouya, B. y Zgaya, H. (2016). "Simulating Complex systems: Complex system theories, their behavioural characteristics and their simulation". En Filipe, J. y van den Herik, J. (eds.), *ICAART 2016: Proceedings of the 8th International Conference on Agents and Artificial Intelligence*, pp. 298-305. Portugal: SCITEPRESS.

Banco Mundial (2019). *Informe sobre el Desarrollo Mundial 2019: la naturaleza cambiante del trabajo*. Washington, DC: Banco Mundial.

Bastian, M.; Heymann, S. y Jacomy, M. (2009). "Gephi: an open source software for exploring and manipulating networks". *ICWSM*, vol. 3, nº 1, pp. 361-362.

Batagelj, V. y Mrvar, A. (1998). "Pajek –a program for large network analysis". *Connections*, vol. 21, nº 2, pp. 47-57.

Bonabeau, E. (2002). "Agent-Based Modeling: Methods and Techniques for Simulating Human Systems". *PNAS*, vol. 99, nº 3, pp. 7280-7287. DOI: https://doi.org/10.1073/pnas.082080899.

Bonacich, P. (1987). "Power and Centrality: A Family of Measures". *American Journal of Sociology*, vol. 92, nº 5, pp. 1170-1182. Disponible en: https://www.jstor.org/stable/2780000.

Bonvecchi, A. y Scartascini, C. (2020). *Who Decides Social Policy?: Social Networks and the Political Economy of Social Policy in Latin*

America and the Caribbean. Washington, DC: Inter-American Development Bank/The World Bank.

Borgatti, S. P.; Brass, D. J.; Labianca, G. y Mehra, A. (2009). "Network Analysis in the Social Sciences". *Science*, vol. 323, nº 5916, pp. 892-895. DOI:10.1126/science.1165821.

Caro, C.; Galindres, D. y Soto, J. (2013). "Sociedad en movimiento: un análisis de redes sociales". *Scientia et Technica*, vol. 18, nº 3, pp. 490-497.

Catebiel, V.; Castro, G. y Hernández, U. (2006). "El Análisis de Redes Sociales en Procesos de Formación Avanzada: el caso de ieRed". *Revista ieRed. Revista Electrónica de la Red de Investigación Educativa*, vol. 1, nº 4, pp. 1-12.

Dahesh, M. B.; Hamidizadeh, M.; Tabarsa, G. y Zandieh, M. (2020). "Reviewing the intellectual structure and evolution of the innovation systems approach: A social network analysis". *Technology in Society*, vol. 63. DOI: 10.1016/j.techsoc.2020.101399.

de la Cruz Flores, G. (2017). "Igualdad y equidad en educación: retos para una América Latina en transición". *Educación*, vol. 26, nº 51, pp. 159-178. Disponible en: http://www.scielo.org.pe/pdf/educ/v26n51/a08v26n51.

Dongsheng, Y. y Yongan, Z. (2008). "Innovation Based on Multi-Agent Method". *2008 International Conference on Computer Science and Software Engineering*, pp 528-531. USA: IEEE.

Epstein, J. M. (2008). "Why Model?". *Journal of Artificial Societies and Social Simulation*, vol. 11, nº 4. Disponible en: http://jasss.soc.surrey.ac.uk/11/4/12.html.

Falco, M.; Giandini, R. y Kuz, A. (2016). "Análisis de redes sociales: un caso práctico". *Computación y Sistemas*, vol. 20, nº 1, pp. 89-106. DOI: 10.13053/CyS-20-1-2321.

Faust, K. y Wasserman, S. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Franco-Bermúdez, J. F. y Ruiz-Castañeda, W. L. (2019). "Análisis de redes sociales para un sistema de innovación generado a partir de un modelo de simulación basado en agentes". *TecnoLógicas*, vol. 22, nº 44, pp. 23-46. DOI: http://dx.doi.org/10.22430/22565337.1183.

Frantz, T. L. (2012). "Advancing Complementary and Alternative Medicine through Social Network Analysis and Agent-Based Modeling". *Forsch Komplementmed*, vol. 19, nº 1, pp. 36-41.

Freeman, L. (1978). "Centrality in social networks conceptual clarification". *Social networks*, vol. 1, nº 3, pp. 215-239.

_____ (2004). *The development of social network analysis*. Vancouver: ⊣P Empirical Press.

García Vázquez, J. y Sancho Caparrini, F. (2016). *NetLogo: una herramienta de modelado*.

Garcia, R. y Jager, W. (2011). "From the Special Issue Editors: Agent-Based Modeling of Innovation Diffusion". *Product Development & Management Association*, vol. 28, nº 2, pp. 148-151.

García-Guadilla, C. (2013). "Universidad, desarrollo y cooperación en la perspectiva de América Latina". *Revista Iberoamericana de Educación Superior*, vol. 4, nº 9, pp. 21-33.

García-Valdecasas, J. (2014). "El impacto de la estructura de las redes sociales sobre el acceso de los individuos al mercado laboral". *Revista Internacional de Sociología*, vol. 72, nº 2, pp. 303-321.

Gaudin, Y. y Pérez, R. (2014). "Science, technology and innovation policies in small and developing economies: The case of Central America". *Research Policy*, vol. 43, nº 4, pp. 749-759.

Ghoshal, S. y Nahapiet, J. (1998). "Social Capital, Intellectual Capital, and the Organizational Advantage". *The Academy of Management Review*, vol. 23, nº 2, pp. 242-266.

Gonzáles-Aguilar, A. y Pinto, A. (2014). "Visibilidad de los estudios en análisis de redes sociales en América del Sur: su evolución y métricas de 1990-2013". *Transinformação*, vol. 26, nº 3, pp. 253-267. DOI: https://doi.org/10.1590/0103-3786201400030003.

Granovetter, M. (1983). "The Strength of Weak Ties: A Network Theory Revisited". *Sociological Theory*, vol. 1, pp. 201-223.

Holland, J. H. (2004). *El orden oculto: de cómo la adaptación crea la complejidad*. México, DF: Fondo de Cultura Económica.

Jianhua, L.; Wenrong, L. y Xiaolong, X. (2008). "Research on Agent-based Simulation Method for Innovation System". *International Confe-*

rence on Information Management, Innovation Management and Industrial Engineering, pp. 431-434. USA: IEEE.

Luke, D. A. (2015). *A user's guide to network analysis in R*. New York: Springer.

Luke, D. A. y Stamatakis, K. A. (2012). "Systems Science Methods in Public Health: Dynamics, Networks, and Agents". *Annual Review of Public Health*, vol. 33, pp. 57-76.

Macal, C. M. (2010). "To Agent-Based Simulation from System Dynamics". En Hugan, J.; Jain, S.; Johansson, B.; Montoya-Torres, J. y Yücesan, E. (eds.), *Proceedings of the 2010 Winter Simulation Conference*, pp. 371-382. Baltimore, Maryland: Winter Simulation Conference.

Macal, C. M. y North, M. J. (2006). "Introduction to Agent-based Modeling and Simulation". USA: Argonne National Laboratory. Disponible en: https://www.mcs.anl.gov/~leyffer/listn/slides-06/MacalNorth.pdf.

Miceli, J. E. (2008). "Los problemas de validez en el análisis de redes sociales: algunas reflexiones integradoras". *Revista hispana para el análisis de redes sociales*, vol. 14, nº 1, pp. 1-45.

Mitchell, M. (2009). *Complexity: A Guided Tour*. Oxford, Reino Unido: Oxford University Press.

Morales, R. y Sifontes, D. (2014). "Las patentes como resultado de la cooperación en I+D en América Latina: hechos y desafíos". *Investigación & Desarrollo*, vol. 22, nº 1, pp. 2-18.

Moriello, S. A. (2013). *Ciencias de la complejidad: una breve introducción*. Buenos Aires: Nueva Librería.

Peyró, B. (2015). "Conectados por redes sociales: introducción al análisis de redes sociales y casos prácticos". *Revista hispana para el análisis de redes sociales*, vol. 26, nº 2, pp. 236-241.

Quezada, A., & Canessa, E. (2010). Modelado basado en agentes: una herramienta para complementar el análisis de fenómenos sociales. *Avances en Psicología Latinoamericana, 28*(2), 226-238.

Rahmandad, H. y Sterman, J. (2008). "Heterogeneity and Network Structure in theDynamics of Diffusion: Comparing Agent-Based and Differential Equation Models". *Management Science*, vol. 54, nº 5, pp. 998-1014.

Ruiz, W. L. (2016). *Análisis del impacto de los intermediarios en los sistemas de innovación: una propuesta desde el modelado basado en agentes*. Medellín: Universidad Nacional de Colombia.

Sanz, L. (2003). "Análisis de redes sociales: o como representar las estructuras sociales subyacentes". *Apuntes de Ciencia y Tecnología*, vol. 7, pp. 21-23.

Sargent, R. G. (2005). "Verification and Validation of Simulation Models". En Armstrong, F. B.; Joines, J. A.; Kuhl, M. E. y Steiger, N. M. (eds.), *Proceedings of the 2005 Winter Simulation Conference*, pp. 130-143. USA: IEEE.

Scott, J. (1988). "Social network analysis". *Sociology*, vol. 22, nº 1, pp. 109-127. DOI: https://doi.org/10.1177/0038038588022001007.

_____ (2000). *Social Network Analysis: A Handbook*. Londres: SAGE.

Valente, M. (2008). *Laboratory for Simulation Develpment - LSD*. Pisa, Italia: Laboratory of Economics and Management. Disponible en: http://www.lem.sssup.it/WPLem/files/2008-12.pdf.

van Eck, N. J. y Waltman, L. (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping". *Scientometrics*, vol. 84, nº 2, pp. 523-538. DOI: 10.1007/s11192-009-0146-3.

Velasco, M. (2016). "Cambio de políticas en América Latina: ampliando el debate". *Revista de Ciencias Sociales*, nº 54, pp. 1-10.

Wilensky, U. (1999). *NetLogo*. Evanston, IL: Center for Connected Learning and Computer Based Modeling/Northwestern University. Disponible en: http://ccl.northwestern.edu/netlogo/.

Williams, R. A. (2022). "Towards an agent-based model using a hybrid conceptual modelling approach: A case study of relationship conflict within large enterprise system implementations". *Journal of Simulation*. DOI: 10.1080/17477778.2022.2122741.

Capítulo 2
El análisis de redes sociales: una herramienta de análisis para entender los patrones de creación y difusión de conocimiento. Su aplicación a colaboraciones científicas y de conocimiento interorganizacionales

Lilia Stubrin, Cecilia Tomassini

Introducción

El objetivo de este capítulo es proveer una introducción a la metodología del análisis de redes sociales (ARS) como herramienta para entender los patrones de creación y difusión de conocimiento. El ARS provee un herramental metodológico propio que resulta útil para visibilizar y entender las relaciones y vínculos entre distintos tipos de agentes o actores, así como un marco analítico que ayuda a comprender cómo se crean estas redes entre actores y qué impacto tiene la estructura de la red en su funcionamiento y accionar. En este capítulo, nos focalizamos en particular en la aplicación del ARS a temáticas que atraen cada vez mayor atención dentro del campo de CTI: el entendimiento de las colaboraciones científicas y las colaboraciones de conocimiento interorganizacionales para la creación y difusión de conocimiento.

Las colaboraciones científicas se han acrecentado notoriamente en las últimas décadas como resultado de varios factores: una mayor interdisciplinariedad científica, el desarrollo de nuevos campos de conocimiento

de mayor complejidad, la disponibilidad de tecnologías de la información y la comunicación (TIC) que facilitan el trabajo colaborativo a distancia, la difusión de prácticas de construcción de conocimiento colaborativo, e incluso la mayor disponibilidad de financiamiento para la ciencia que promueve la cooperación internacional, interdisciplinaria e interinstitucional (Gibbons *et al.*, 1994; Gibbons, Ziman, 2000; Gibboms, Nowotny y Scott, 2001; Perez Beltran, Rodriguez-Aceves y Valerio Urena, 2015).

Las colaboraciones de conocimiento interorganizacionales en las que diferentes organizaciones (empresas, universidades, laboratorios y otros actores) participan intercambiando, transfiriendo y cocreando conocimiento también se han venido expandiendo notoriamente en las últimas décadas. Este crecimiento se explica, en gran parte, por la emergencia de nuevas actividades intensivas en conocimiento (como la biotecnología, la nanotecnología o las TIC), en las que las bases de conocimiento son complejas y en constante crecimiento, lo que motiva a las organizaciones a participar de complejos entramados de vinculaciones de manera tal de acceder a nuevo conocimiento y activos complementarios, así como compartir los altos riesgos y costos que conllevan los procesos de creación de conocimiento en ese tipo de actividades (Hagedoorn y Schakenraad, 1992; Eisenhardt y Schoonhoven, 1996; Mowery, Oxley y Silverman, 1998).

El capítulo explora y analiza cómo y de qué manera en el campo de los estudios de CTI se ha utilizado el marco metodológico de ARS para entender cómo se conforman, qué características y estructuras tienen las colaboraciones científicas e interorganizacionales y cómo estas afectan las formas de creación y difusión de conocimiento. El capítulo se estructura de la siguiente manera: luego de esta introducción, en el segundo apartado desarrollamos brevemente el marco analítico del ARS y después introducimos los principales elementos metodológicos del ARS. En el cuarto apartado nos detenemos en la aplicación del ARS para el análisis de las colaboraciones científicas y las colaboraciones de conocimiento interorganizacionales y, finalmente, a modo de reflexión final, destacamos las oportunidades y limitaciones que presenta el ARS como herramienta metodológica para estudiar los procesos de CTI en América Latina y el Caribe.

Marco analítico del ARS y los estudios de CTI

El ARS surge como un marco analítico cuyo origen se remonta a la década del treinta. Este fue evolucionando a partir de aportes de disciplinas tanto ligadas a las ciencias naturales (CCNN) como a las ciencias sociales (CCSS) (Freeman, 2012; Hidalgo, 2016). Mientras que las contribuciones desde la física y la matemática han aportado al conocimiento de las propiedades de las estructuras de las redes, la sociología, la psicología y la economía han contribuido a comprender cómo los comportamientos de los actores pueden ser influenciados y condicionados por la estructura de relaciones en la que están inmersos (Moreno, 1953; Lazarsfeld y Merton, 1954; Granovetter, 1973; White, 1992; Burt, 1995) (ver el siguiente apartado titulado "¿Qué es una red?"). Cabe destacar, sin embargo, que la división disciplinar que tendió a primar en un principio ha tendido a borrarse cada vez más haciendo del ARS un área de conocimiento que se ha ido consolidando como interdisciplinaria, en la cual las CCSS y las CCNN dialogan y se integran dentro de un paradigma organizado. En las últimas décadas, además, el uso del ARS como una herramienta metodológica de análisis empírico de redes se ha extendido notoriamente asociado a la disponibilidad de datos a gran escala y a la emergencia y difusión de software especializados.

En el campo específico de los estudios de CTI, el ARS ha contribuido con elementos conceptuales al entendimiento de cómo se forman o cuál es el origen de las redes de creación o difusión de conocimiento, y cómo la estructura de esas redes afecta su funcionamiento y el accionar individual de los nodos que la componen.

En cuanto al origen de las redes, los aportes teóricos han buscado entender las razones por las cuales los vínculos o lazos son creados entre actores y cómo al formarse influyen en la conformación de la estructura de la red en términos de, por ejemplo, conectividad, centralidad de determinados nodos o formación de grupos o clústers (Borgatti *et al*., 2009). En el campo de la CTI la investigación sobre conformación y origen de las redes de conocimiento busca responder preguntas del tipo: ¿qué factores explican la creación de redes de investigación científica para la creación de nuevo conocimiento? ¿Qué razones motivan a las empresas a establecer colaboraciones de I+D con universidades? ¿Qué factores influyen en la elección de ciertas organizaciones sobre otras? Las hipótesis dominantes en la literatura respecto a cómo eligen los actores de una red con quién o quiénes establecer vínculos de conocimiento están asociadas a tener

rasgos en común (tamaño, sector, capacidades tecnológicas, etc.), compartir lazos con los mismos nodos ("los amigos de mis amigos, son mis amigos"), compartir una historia común (ya haber establecido vínculos en el pasado) y aliarse estratégicamente para acceder a algún recurso o a alguna capacidad (Lazarsfeld y Merton, 1954; Granovetter, 1973, 1985; Borgatti *et al*., 2009; Jackson, 2010; Freeman, 2012).

Otra área de estudio dentro del ARS está ligada al análisis de las estructuras de las redes. Las redes, en su condición de conjunto de actores y relaciones, según cómo se establezcan los lazos entre los actores, pueden adoptar distintas estructuras. Estas estructuras ("quién se conecta con quién") pueden tener implicancias tanto para la acción individual (la capacidad de innovación de una empresa, el acceso a conocimiento diverso, entre otros) como para la distribución de recursos dentro de la red (por ejemplo, la difusión de conocimiento dentro de una comunidad científica o en una industria) (Faust y Wasserman, 1994; Ahuja, 2000; Cowan, 2004; Cowan y Jonard, 2004, 2007, 2008). Cabe destacar que no hay estructuras buenas o malas *per se*, sino que algunas estructuras son más favorables que otras para ciertos objetivos de la red. Por ejemplo, redes con estructuras cerradas, es decir, con altos niveles de conectividad entre los nodos, y no jerárquicas se ha encontrado que favorecen la creación de capital social a través de la emergencia de lazos de confianza entre los nodos y la minimización de conductas oportunistas. Este tipo de estructura de red se ha encontrado como muy favorable para la difusión y creación de conocimiento entre organizaciones (Coleman, 1988; Koput, Powell y Smith-Doerr, 1996, 1997; Gargiulo y Gulati, 1999; Dyer y Nobeoka, 2000). Otro tipo de estructuras, como aquellas con lazos que unen dos redes desconectadas ("agujeros estructurales") otorgan más oportunidades para controlar, distribuir y diversificar conocimiento que circula entre grupos opuestos y distantes (Burt, 2004).

Una vertiente de la literatura se ha dedicado a la modelización y teorización de las estructuras de redes. Las estructuras más estudiadas se denominan aleatoria (Erdős y Rényi, 1961), de mundo pequeño (Strogatz y Watts, 1998) y libre de escala (Albert y Barabási, 1999). La red de mundo pequeño, por ejemplo, con una estructura en la que los nodos tienden a estar muy conectados entre ellos ("mis amigos son amigos entre ellos"), pero cada nodo tiene lazos con otros fuera de su entorno cercano, ha

sido encontrada empíricamente en un número importante de redes.[2] Las razones por las cuales las redes con este tipo de estructura emergen con frecuencia en la realidad no son *a priori* evidentes, sin embargo, gran parte de la literatura apela a explicaciones vinculadas a la creación de capital social y confianza. Estos últimos son especialmente importantes, por ejemplo, cuando las colaboraciones involucran intercambiar conocimiento con terceros, como es el caso de las alianzas de I+D (Koput, Powell y Smith-Doerr, 1996, 1997; Gargiulo y Gulati, 1999; Dyer y Nobeoka, 2000) o de las redes de colaboración científica (Newman, 2001; Kastelle y Steen, 2010), por ejemplo.

En suma, distintos aportes, algunos de los cuales fueron mencionados en esta sección, han contribuido a generar un marco conceptual asociado al ARS crecientemente utilizado para estilizar conceptos de la teoría evolucionista y entender desde una óptica formal los procesos de creación y difusión de conocimiento.

¿Qué es una red?

Una red se define como un grupo de nodos (o actores) unidos por lazos (o vínculos). Casi cualquier fenómeno puede pensarse como una red. Los nodos o actores que componen una red pueden ser individuos, grupos de personas, empresas, investigadores o países, entre otros. Los vínculos son en general flujos tangibles o intangibles que conectan los nodos como, por ejemplo, transferencias monetarias, amistad, acuerdos de investigación y desarrollo, flujos de comercio, información, transacciones comerciales, entre otros (Faust y Wasserman, 1994; Durrance y Williams, 2008). En el ARS, la unidad de análisis no son los nodos, sino el conjunto formado por ellos y sus vínculos, de manera tal que se considera que: 1) los actores y sus acciones son interdependientes (no autónomos o independientes), 2) los lazos entre los actores transfieren recursos (materiales o simbólicos) y 3) las estructuras de relaciones influyen (favoreciendo y/o obstaculizando) en la acción individual (Faust y Wasserman, 1994).

2 Por ejemplo, la red de actores de Hollywood (Strogatz y Watts, 1998), la red de alianzas estratégicas entre empresas en varias industrias (Ahuja, 2000; Pammolli y Riccaboni, 2002; Baum, Rowley y Shipilov, 2003; Duysters y Verspagen, 2004; Koput *et al*., 2005) o las redes de coautorías de trabajos científicos (Barabási, 2002).

Aproximaciones metodológicas al ARS

En esta sección abordamos los principales elementos metodológicos del ARS. En la primera subsección, el análisis se concentra en la definición de la unidad y el nivel de análisis, mientras que en la segunda subsección se abordan los requerimientos de datos y los principales indicadores para el ARS.

Definición de unidad y nivel de análisis

Formalmente, una red *g* está compuesta por *N*, $N = (1, 2,... ,n)$, nodos y las relaciones que existen entre ellos. Si $g_{ij} = 1$ existe un lazo entre los nodos *i* y *j* en la red *g*. Si $g_{ij} = 0$, no hay lazo que vincule a *i* con j en la red *g*. Dependiendo del tipo de relación que componga la red, una red puede tener lazos no direccionados (o simétricos) o lazos dirigidos (o asimétricos). Las relaciones de amistad, de parentesco, de acuerdos de I+D entre empresas son ejemplos de lazos no direccionados. En este tipo de vínculos no hay origen ni destino de la relación, de manera tal que, si *x* tiene un lazo con *y*, entonces, necesariamente *y* tiene un lazo con *x*. En cambio, los vínculos dirigidos son, entre otras, las asistencias tecnológicas (que pueden ser brindadas o recibidas), las citas bibliográficas, la otorgación de licencias o las deudas. Este tipo de lazos tiene una direccionalidad. En el caso de las citas bibliográficas que "el autor *x* cite al autor *y*" no implica que "el autor *y* cite al autor *x*". En muchos casos, importa no solo la existencia de los vínculos sino su intensidad para comprender la conformación y la dinámica de la red. Por ejemplo, si la empresa A ha tenido cinco acuerdos de I+D con la empresa B en los últimos tres años, A y B están conectadas por un vínculo no dirigido, y el valor de ese vínculo es mayor que en el caso de firmas que hayan firmado tan solo un acuerdo de I+D en el mismo período. En algunas aplicaciones reconocer el valor o la fuerza del vínculo puede ser relevante. Tal es el caso, por ejemplo, del valor de las transacciones comerciales entre países, si se quiere entender mejor la participación de los países en el flujo de comercio internacional, o la frecuencia de interacciones cara a cara entre investigadores, si se busca comprender la cercanía del vínculo entre estos. En este tipo de casos en que se mide la intensidad del vínculo, se dice que los vínculos son ponderados y la red se denomina "red ponderada".

El análisis de una red puede realizarse en distintos niveles analíticos. En un extremo, el análisis es de "red completa" cuando se cuenta con datos sobre los N actores de la red. En el otro extremo, el análisis puede ser "egocéntrico" tomando cada nodo como unidad de análisis. En este último caso, si una red tiene N nodos, el estudio tiene N unidades de análisis.[3] Otros niveles de análisis, aunque menos frecuentes en la literatura, son los pares de nodos (*dyadic network*) o conexiones entre tres nodos (*triadic network*). La elección del nivel de análisis depende del conocimiento previo que se tenga de la población de nodos de la red, de la disponibilidad y accesibilidad de los datos sobre los nodos y del objeto de estudio. En la próxima sección proveemos ejemplos de distintos niveles de análisis para el estudio de colaboraciones científicas y alianzas interorganizacionales.

Datos y medición

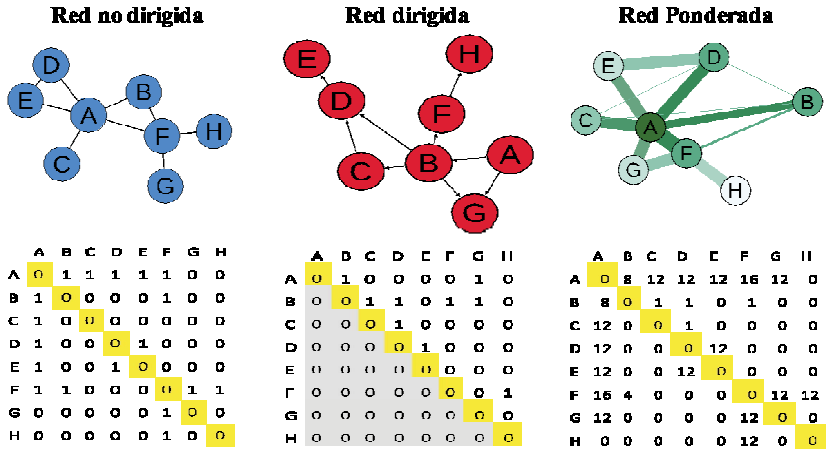
La recolección de datos es un aspecto metodológico crucial en el ARS. Los datos pueden tener dos orígenes: 1) datos secundarios o prefabricados en registros como, por ejemplo, patentes, publicaciones académicas y datos de comercio mundial y 2) datos primarios generados a través de relevamientos *ad hoc*, como encuestas. En el primer caso, la recolección de datos implica explorar la existencia y accesibilidad de fuentes de información para construir datos relaciones, es decir, compuestos por los nodos y sus vínculos de interés (es decir, para extraer datos de vinculaciones entre investigadores por ejemplo a través de la coautoría de artículos científicos). En el segundo caso, la recolección de datos requiere del diseño de un instrumento (es decir, cuestionario) que permita recabar datos sobre los nodos y los vínculos de la red. Ejemplos de la aplicación de ambos tipos de relevamientos de datos en el caso de estudios de CTI se discuten en la siguiente sección.

La organización de los datos de redes para su análisis es también central. En el ARS lo que importa es la construcción de datos relacionales para lo que se utiliza una matriz de adyacencia. Las columnas y filas de la matriz representan los nodos (actores), mientras que las celdas de la matriz indican la presencia (1) o ausencia (0) de relaciones entre pares de nodos, o la frecuencia/importancia de vínculo entre actores. La figura 1

3 Para profundizar en el nivel de análisis de las redes ego se recomienda la lectura de Molina González (2005) y Borgatti, Perry y Pescosolido (2018).

muestra tres ejemplos de matrices de adyacencia y su visualización en redes según el tipo de vínculo (no dirigido, dirigido o ponderado). La representación gráfica (o grafos) de estas matrices de relaciones se utiliza frecuentemente como una herramienta para describir las redes y facilitar la identificación de patrones, estructuras o características de las redes. La ubicación de nodos y lazos en el espacio del gráfico puede tomar diferentes formas, pero lo importante es el patrón de conexiones y no la posición actual en el espacio. Cuando dos puntos están conectados se dice que los nodos son adyacentes, las conexiones recíprocas se representan gráficamente con líneas y las conexiones dirigidas con flechas que indican la dirección de la relación.

Figura 1. Redes según tipos de vínculos y matrices de adyacencia



Fuente: elaboración propia, software Gephi

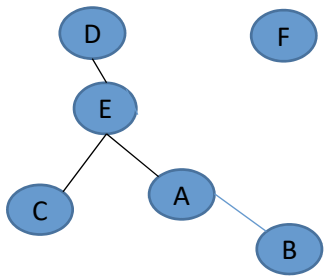
El análisis de los datos se realiza utilizando indicadores, los que se pueden agrupar en dos tipos: 1) indicadores sobre la estructura general de la red, o la topología y 2) indicadores sobre los componentes individuales de la red y sus atributos.[4]

En cuanto a los indicadores sobre la estructura general de la red (ver cuadro 1), entre los más básicos se encuentra el tamaño de la red

4 En este apartado nos basamos principalmente en los siguientes manuales metodológico: Faust y Wasserman (1994) y Borgatti, Everett y Johnson (2013).

(el número total de nodos y de vínculos). La red de la figura 2 tiene seis nodos y cinco vínculos. En las redes los nodos pueden o no estar contactados. Si lo están, pueden estar conectados directa o indirectamente. En la figura 2 los nodos A y B están directamente conectados, mientras que los nodos B y E están indirectamente conectados. Los nodos pueden estar indirectamente conectados por dos o más nodos. Se llama camino la sucesión finita de lazos que conecta a cualesquiera dos nodos *i* y *j*. La longitud del camino está determinada por la cantidad de lazos que lo componen. La *distancia* o *longitud* promedio entre nodos de una red mide la separación entre cualesquiera dos nodos. En la figura 2, los nodos B y E están conectados por un camino de longitud 2. En cambio, los nodos B y D tienen más de un camino que los conecta. El camino con longitud más corta se denomina camino geodésico (dos nodos pueden estar conectados por más de un camino geodésico). Cuando en un camino no se repite ningún lazo ni ningún nodo, este se denomina recorrido. Y los recorridos que comienzan y terminan en el mismo nodo, son denominados ciclos. En la figura 2 hay un ciclo compuesto por los nodos A, C y E.

Figura 2. Caminos, recorridos y ciclos



Fuente: elaboración propia, software Gephi

En algunos casos, como en la figura 2, la red no está completamente conectada. Esta tiene más de un componente, un componente principal (llamado componente gigante) y un componente secundario, que en este

caso es un nodo aislado.[5] Un componente es el conjunto máximo de nodos en el que cada nodo puede alcanzar cualquier otro a través de algún camino (sin importar cuán largo sea este).

Otras de las medidas importantes para caracterizar la estructura de las redes puede ser la de destacar el diámetro y la densidad. En el primer caso, el diámetro se refiere a la máxima distancia entre cualesquiera dos nodos de una red. El diámetro de una red tiene un mínimo de 1 (si una red es completa) y un máximo de N-1 (o infinito en el caso de que se trate de una red no conectada). Cuando las redes tienen más de un componente, el cálculo del diámetro de la red se realiza para el componente gigante, es decir, el más grande de la red. Mientras que la densidad mide qué tan cerca está la red de estar completa. Es la proporción de vínculos presentes sobre los posibles. Una red completa significa que todas las conexiones posibles tienen lugar entre todos los nodos, en este caso la densidad es 1. Un problema de esta métrica es que no se puede utilizar para la comparación entre redes de diferentes tamaños, ya que el denominador depende de la cantidad de nodos de cada red. Alternativamente, se utiliza el grado promedio de la red, que es simplemente el promedio del total de conexiones de cada nodo de la red.

Los indicadores de agrupamientos son también importantes y frecuentemente utilizados para entender la estructura de las redes pues indican la conformación de grupos dentro de la red global, es decir, indican cuán conectados están los nodos que la componen.[6] El coeficiente de agrupamiento (CA) mide el grado en que una red no dirigida tiene áreas de alta y baja densidad.[7] A nivel del nodo, el coeficiente de agrupamiento individual (CAI) mide la densidad de vínculos en la red de cada nodo (es decir, la densidad de vínculos entre los nodos conectados a un nodo dado). Es decir que, si el nodo *i* tiene un lazo con *j* y otro lazo con *k*, el grado de agrupamiento mide si *k* y *j* están, a su vez, conectados entre ellos. Por lo tanto, el coeficiente de agrupamiento de un nodo *i* es la cantidad de veces que ocurre que sus vínculos están conectados entre ellos (es decir, "que tan amigos son los amigos de *i* entre sí"). El cálculo del coeficiente de agrupamiento del nodo *i* consiste en medir todos los pares de nodos que están conectados directamente con *i*, y luego calcular cuántos de estos están además conectados entre sí. Luego para el cálculo

5 La definición de *componente* utilizada permite considerar un nodo aislado (sin vínculos) como un *componente* en sí mismo.

6 Esta característica se conoce también como transitividad, *cliquishness* o *clustering*.

7 Este indicador es útil para determinar si se trata de redes de mundo pequeño, ver el capítulo 14 de Borgatti, Everett y Johnson (2013).

del CA global de la red se promedia esta cantidad en todos los nodos. A partir de los coeficientes de agrupamiento para cada nodo de la red se puede calcular el grado de agrupamiento promedio de la red.

Los métodos para la detección de comunidades dentro de las redes es uno de los temas que ha recibido mayor atención en los últimos tiempos (Newman, 2006). Estos métodos normalmente asumen que la red se divide naturalmente en subgrupos, en los que el número y tamaño de los grupos es determinado por la propia red y no *a priori* por el investigador. Existe una gran diversidad de métodos para identificar comunidades. La modularidad es definida por Newman como la fracción de los bordes que caen dentro de un grupo, menos el valor esperado que tendría esta fracción si los enlaces se distribuyeran aleatoriamente. La modularidad varía entre -1 y 1, los valores cercanos a 1 indican una alta modularidad. Las redes altamente modulares tienen conexiones sólidas entre los nodos que componen una comunidad y conexiones débiles entre los nodos en diferentes comunidades. Esto significa que algo diferente (por ejemplo, mayores niveles de interacción, intercambio de información, etc.) está sucediendo entre ciertos grupos en comparación con otros.

Cuadro 1. Indicadores sobre la estructura general de la red

Indicador	Notación y cálculo	Interpretación
Longitud	$L = \frac{1}{N}\sum_{i}\sum_{j \neq i}\frac{d(i,j)}{N-1}$ En la que $d(i,j)$ es la distancia geodésica entre i y j	Mide la separación entre cualesquiera dos nodos.
Densidad	$D = 2\ \frac{t}{n(n-1)}$ En una red no dirigida	Mide qué tan cerca está la red de estar completa, es decir, de expresar todas las conexiones posibles entre sus nodos.
Coeficiente de agrupamiento	$C = \frac{1}{N}\sum_{i}\sum_{(j,k \in \Gamma_i)}\frac{X(j,k)}{\|\Gamma_i\|\,(\|\Gamma_j\|-1)/2}$ En la que $X(j,k) = 1$ si $j \in \Gamma$, $X(j,k) = 0$, caso contrario, Γi son los nodos directamente conectados a i, y $\|\Gamma i\|$ es la cantidad de nodos directamente conectados a i	Mide la conformación de grupos dentro de la red global, es decir, cuán conectados están los nodos que la componen.

Fuente: elaboración propia

En cuanto a las métricas para estudiar los componentes individuales de la red, se destacan las medidas de centralidad (cuadro 2). Estas miden la importancia o prominencia de un nodo en la red. En la literatura se utilizan fundamentalmente tres tipos de centralidad: de grado, de intermediación y de cercanía. La centralidad de grado se define como el número de actores con los que cada nodo se vincula directamente (o el número de nodos adyacentes). En el caso de redes dirigidas, se diferencia entre el grado de salida (número de nodos adyacentes con los que el nodo *x* se vincula) y el grado de entrada (número de nodos adyacentes que se vinculan con el nodo *x*). Dependiendo del tipo de flujo de la red, el grado de salida se puede interpretar como la influencia de un actor, mientras que el grado de entrada podría interpretarse como prestigio o popularidad del actor. La centralidad de grado no distingue qué tan conectados o centrales son los nodos adyacentes. Estos podrían ser nodos que no tengan otras conexiones o incluso los nodos más populares de la red. Una variación de la centralidad de grado que pondera cada nodo adyacente por su centralidad es la centralidad de vector propio (CVP) (Bonacich, 1972). Dependiendo del tipo de flujos de la red, la CVP puede interpretarse, por ejemplo, como una medida de popularidad, un nodo con alta CVP está conectado a nodos que están bien conectados.

La centralidad de la intermediación (Freeman, 1979) indica la frecuencia con la que un nodo está en el camino más corto entre otros pares de nodos. Esto en el entendimiento de que un actor que se coloca entre las vías de comunicación de los demás tiene una centralidad derivada del potencial de control de los flujos y comunicaciones de la red. Por último, la centralidad de cercanía da cuenta de qué tan cerca está un nodo de todos los demás en la red. Esta es la suma de las distancias geodésicas (longitud del camino más corto) que conecta un nodo a todos los demás. En la versión normalizada del indicador (en la que la centralidad de cada nodo se divide en *n -1*) valores cercanos a 0 indican que se trata de un nodo periférico en la red, mientras que valores cercanos a 1 indican que es un nodo más central. En una red de flujos de información, por ejemplo, la centralidad de cercanía es útil para interpretar quién accede más rápido a cierta información. Un nodo con alta CC puede acceder antes a la información que surge de cualquier nodo aleatorio de la red, mientras que un nodo periférico demorara más tiempo y es probable que reciba la información mayor distorsión.

Cuadro 2. Medidas de centralidad

Medidas de centralidad*	Notación y cálculo	Interpretación
Centralidad de grado (G)	$G_i = \sum_j xij$ En una matriz de red no dirigida, la centralidad de grado del actor *i (Gi)* y del actor *xij* es la entrada *(i, j)* de la matriz de adyacencia.	Número de vínculos directos que tiene un nodo o número de nodos adyacentes. Se puede calcular como un indicador ponderado por el peso del vínculo.
Centralidad de vector propio (CVP)	$CVP_i = \lambda \sum_j x_{ij} e_j$ En la que la CVP es la puntuación de centralidad del vector propio; λ es una constante de proporcionalidad llamada valor propio. La centralidad de cada nodo es proporcional a la suma de centralidades de los nodos a los que es adyacente.	Variación de la centralidad de grado que pondera cada nodo adyacente por su centralidad.
Centralidad de intermediación (CI)	$CI_j = \sum_{i<k} \frac{g_{ijk}}{g_{ik}}$ En la que *gijk* es el número de caminos geodésicos que conectan *i* y *k* a través de *j*, y *gik* es el número total de trayectorias geodésicas que conectan *i* y *k*. La CI de un nodo es cero cuando nunca se encuentra en el camino más corto entre otros dos. El mayor valor de CI se da cuando el nodo se encuentra a lo largo de cada camino más corto entre cada par de otros nodos.	Mide qué tan a menudo un nodo determinado se encuentra en el camino más corto entre otros dos nodos.
Centralidad de cercanía (CC)	$CC = \frac{1}{(\sum_y d(y,x))}$ En la que *d (y, x)* es la distancia geodésica entre los nodos *y* y *x*.	Es una medida de qué tan cerca está un nodo de todos los demás en la red, es decir, la distancia total (en el gráfico) de un nodo dado de todos los demás nodos.

Fuente: elaboración propia

Nota: (*) Para redes no dirigidas.

En la práctica, diversos tipos de software se utilizan para el procesamiento, análisis y visualización de los datos de redes.

Softwares recomendados para el ARS y su visualización

Existe una amplia disponibilidad de software para el ARS. Su selección para el análisis dependerá de los objetivos de este. Por ejemplo, Gephi y R/igraph son recomendados para trabajos enfocados en la visualización de las redes y el uso de grandes bancos de datos. En cambio, Pajek, PNet o R/igraph se recomiendan más para el modelado de datos de red.

Cuadro 3. Softwares

Softwares	Sitios web	Manuales
UCINET	https://sites.google.com/site/ucinetsoftware/home	Borgatti, Everett y Johnson (2013)
Gephi	https://gephi.org/	Cherven (2015)
Igraph	https://igraph.org/	Luke (2015)
Pajek	http://mrvar.fdv.uni-lj.si/pajek/	Batagelj, De Nooy y Mrvar (2018)
PNet	http://www.melnet.org.au/pnet	Koskinen, Lusher y Robins (2012)

Fuente: elaboración propia

Aplicación del ARS en estudios de CTI

En esta sección exploramos cómo el ARS es aplicado para el estudio de dos fenómenos: las redes de colaboración científica y las redes de conocimiento interorganizacionales. Las preguntas que guían la sección, y que son respondidas para cada caso, son: ¿cuáles son las unidades y niveles de análisis? ¿Qué tipo de vínculos se estudian? ¿Qué tipo de indicadores se utilizan? ¿Cuáles son las principales fuentes de datos utilizadas? ¿Qué aspectos metodológicos y conceptuales hay que tener en cuenta para la realización de este tipo de estudios en América Latina y el Caribe?

Este apartado está divido en tres subapartados. En el primero y segundo apartados se analizan los aspectos metodológicos para el análisis de redes de colaboración científica y redes de conocimiento interorganizacionales, respectivamente. En cada uno de estos subapartados el análisis se organiza a partir de la definición de la unidad y el nivel de análisis y la medición y obtención de datos para el ARS. Finalmente, se provee de un diagrama a modo de resumen de la aplicación del método de ARS al análisis de las redes de conocimiento.

Redes de colaboraciones científica

Definición de unidad y nivel de análisis

El estudio de las colaboraciones científicas mediante el ARS puede ser sistematizado a partir de diversos recortes o niveles, por ejemplo: 1) individual, 2) institucional u organizacional, 3) internacional, regional o local o 4) cognitivo.

Los estudios que se enfocan en el nivel individual en general toman como unidades de análisis a los investigadores que colaboran entre sí, o con actores por fuera del ámbito científico. Su análisis se puede abordar desde indicadores de la estructura de las redes, revelando por ejemplo cuán fragmentada o integrada es una comunidad de científicos, o desde indicadores a nivel de nodo para identificar quiénes son los principales actores de la red o qué investigadores están mejor conectados para acceder a ciertos recursos. Varios autores, en especial desde la bibliometría, se han preguntado cómo la localización de los autores en ciertas redes puede explicar resultados y desempeños en términos de productividad, calidad de la investigación o el impacto y visibilidad de las publicaciones (Beaver, 2013). Se estudia, por ejemplo, el efecto de diversas medidas de centralidad en redes de coautoría sobre el impacto de las publicaciones (medido como citaciones) (Li, Liao y Yen, 2013), o la influencia de la estructura de redes (densidad e integración) sobre la productividad y el impacto (González-Brambila, Krackhardt y Veloso, 2013). Otros estudios están interesados en explicar cómo ciertos atributos personales determinan las posibilidades de colaboración y la posición de los investigadores en redes científicas. Por ejemplo, se ha estudiado cómo el género puede jugar un papel en la formación de redes de coautoría, o si la posición de las mujeres científicas en las redes de colaboración tiene impactos en su productividad y en la ampliación de las brechas de género (Araújo y Fontainha, 2017).

En el nivel institucional u organizacional en general encontramos trabajos que definen sus unidades de análisis a partir de los agregados de pertenencia de los individuos (departamentos de investigación, instituciones, laboratorios, empresas, etc.). El ARS se utiliza en general para estudiar las colaboraciones dentro de las propias instituciones (intrainstitucionales), entre instituciones diferentes y según el tipo de institución (interinstitucionales), para entender el funcionamiento de determinados sectores, clústers o áreas específicas. Por ejemplo, dada la alta dispersión

de los conocimientos y tecnologías en ciencias médicas se ha prestado especial atención sobre cómo la división del trabajo en redes intra e interorganizacionales influye en su evolución y trayectorias (Consoli y Ramlogan, 2008, 2012; Consoli y Mina, 2009). Algunas de las preguntas que estos trabajos buscan responder son ¿qué tipos de organizaciones participan en la producción de conocimiento en ciencias médicas? ¿Cómo cambia su contribución con el tiempo? ¿Cuál es la correspondencia entre el tipo de organización y el área de especialización? ¿Cómo emergen y evolucionan los patrones de desarrollo del conocimiento científico en redes de colaboración en esta área? ¿Cómo se refleja la complejidad de los problemas clínicos de salud en los sistemas de innovación a partir de redes interinstitucionales de colaboración? El uso de ARS en este nivel ha sido clave también para el estudio de la producción de conocimientos, tecnologías e innovaciones movilizada por las relaciones universidad-empresa o universidad-industria. En América Latina son varios los trabajos que exploran las redes interinstitucionales de colaboración científica en el área de salud, mostrando entre otras cosas la relevancia de las instituciones públicas y la falta de articulación con actores del sector productivo o empresarial (de Paula Fonseca e Fonseca, Fernandes y Fonseca, 2017; Cohanoff *et al*., 2021; dos Reis Azevedo Botelho *et al*., 2022).

Los estudios de red a nivel internacional o regional por lo general definen sus unidades de análisis a partir de la agregación geográfica de investigadores individuales o instituciones según países, regiones o en función de recortes territoriales locales dentro de los países (municipios, estados, comunidades, etc.) (Katz, 1994; Duque *et al*., 2005; Leydesdorff y Wagner, 2005; Ainsworth *et al*., 2007; Adshead y Quayle, 2018). Por ejemplo, Costa, da Silva Pedro y de Macedo (2013) analiza las colaboraciones científicas en biotecnología en la región noreste de Brasil a partir de la coautoría entre instituciones, observando la mayor relevancia de los intercambios a nivel intrainstitucional y regional, en detrimento de las colaboraciones interinstitucionales e internacionales. Por su parte de Paula Fonseca e Fonseca *et al*. (2019) observa una escaza colaboración en redes coautoría sur-sur de investigaciones estratégicas como las orientadas a la prevención y/o tratamiento del sida. Otra forma de definición de unidades y vínculos es a partir de diversas actividades de cooperación internacional, como la firma de acuerdos, convenios, transferencias, programas, proyectos conjuntos, etc.

Por último, en el nivel cognitivo, agrupamos aquellos estudios de ARS que analizan las estructuras cognitivas subyacentes a redes académicas

en ciertas áreas o para ciertos problemas de investigación. En estos casos las unidades de análisis pueden ser las áreas de conocimiento o disciplinas de los investigadores, las palabras claves en títulos o resúmenes en las producciones bibliográficas, entre otras. Por ejemplo, Leydesdorff y Rafols (2009) y Leydesdorff, Porter y Rafols (2010) proponen la visualización de redes como herramienta para mapear las transformaciones sociocognitivas de los sistemas de ciencia y tecnología, explorando las colaboraciones a partir de superponer datos de universidades, corporaciones, agencias de financiamiento y temas de investigación. En América Latina este tipo de técnicas ha servido para evidenciar vacíos, sesgos y potencialidades de las agendas de investigación en áreas particularmente relevantes como las enfermedades olvidadas o las enfermedades tropicales (Argote, Collazo y Rivero, 2017; de Paula Fonseca e Fonseca, Fernandes y Fonseca, 2017; de Paula Fonseca e Fonseca *et al*., 2019).

En todos los niveles presentados, el ARS se muestra como una herramienta útil para la generación de evidencia que informe a las políticas de producción de conocimiento y tecnología en América Latina. En particular porque, combinada con otros indicadores, permite conjugar los niveles de análisis individuales y relacionales con los contextos de inserción y aplicación de las redes.

Medición y datos

El estudio de colaboraciones científicas a partir del ARS utilizando vínculos de coautoría es dominante en la literatura. Esto se explica fundamentalmente por la disponibilidad de datos comparables y su accesibilidad (Beaver, 2001). En las últimas décadas han aumentado considerablemente en cantidad y calidad los bancos de datos que registran publicaciones, siendo Web of Science, Scopus y Medline los más difundidos y utilizados. Esta disponibilidad de información estandarizada ha llevado a que el análisis de las redes de coautoría sea un proceso fiable y verificable, que permite llevar a cabo estudios de colaboración para todos los niveles y unidades de análisis mencionados anteriormente. Por ejemplo, algunos de los elementos que estos bancos permiten extraer son: los nombres de los autores, forma de citación, adscripción institucional, dirección institucional, localización geográfica (ciudad, país), tipo de documento, año de publicación, título del artículo, resumen, palabras claves propuestas por el autor y estandarizadas según

la fuente, DOI, revista en la que se publica, idioma del artículo, año de publicación, acceso abierto, fuente de financiación, así como diversas referencias bibliográficas (artículos citados, por ejemplo) e indicadores bibliográficos del propio artículo.

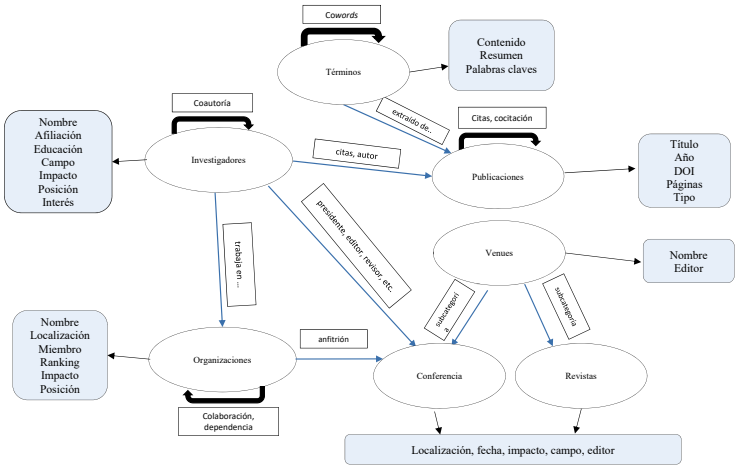
Dentro de la literatura en general, las coautorías se tratan como un sinónimo de la colaboración científica. Sin embargo, esta es apenas una porción mensurable de las muy diversas formas a través de las cuales los científicos comparten sus esfuerzos para producir conocimiento. Entre las limitantes de este tipo de vínculos se señala que no todos los involucrados en una colaboración aparecen representados en las coautorías, surgen así los problemas de subautoría e invisibilización del trabajo de ciertos grupos (Laudel, 2001; Cronin, La Barre y Shaw, 2003). Por el contrario, también puede suceder que autores que firman artículos no hayan participado realmente en la colaboración, mostrando problemas de sobredeclaración y jerarquía (Katz y Martin, 1997; Laudel, 2002). Una desventaja adicional de los estudios de coautorías en artículos científicos surge del sesgo para captar formas más amplias de producción de conocimiento por disciplinas. Esto es especialmente perjudicial para las ciencias sociales (Hicks, 2004), las humanidades, las tecnologías y las ciencias agrarias. Asimismo, se observan sesgos geográficos e idiomáticos (Leta, Sorenson y Vasconcelos, 2009; Chavarro, Rafols y Tang, 2017). Entre las desventajas puntuales del trabajo con estos datos se puede señalar la falta de homogeneización de ciertas categorías (como las instituciones o las direcciones institucionales). Asimismo, cuando se trabaja en niveles de agregación superiores al individual (institucional o por países) los datos de autores con múltiples filiaciones o localizaciones pueden ser problemáticos al sumar colaboración erróneamente (Katz y Martin, 1997).

Además de las redes de coautoría, las redes de citaciones, el acoplamiento bibliográfico y los análisis según el texto (como co*words*) son también usados desde el ARS para estudiar la producción de conocimiento y las colaboraciones. El diagrama 1 resume algunas de las otras unidades utilizadas en el ARS, sus relaciones y atributos (Kong *et al*., 2019).

En la actualidad, varios autores han mostrado la importancia de diversificar las fuentes de información para el estudio de las colaboraciones científicas, en particular aprovechando los avances en la generación de macrodatos académicos en internet y en diversos repositorios nacionales e institucionales. Kong *et al*. (2019) realizan una revisión

de estas fuentes y concluyen que es factible el estudio de las colaboraciones a partir de redes multinivel que permitan integrar una mayor diversidad de unidades de análisis como informes técnicos, propuestas de proyectos, resultados de investigación en informes, materiales educativos, entre otros.

Diagrama 1. Unidades de análisis utilizadas para estudiar colaboraciones científicas



Fuente: Kong *et al.* (2019). Traducción propia

Redes de conocimiento interorganizacionales

Definición de unidad y nivel de análisis

La literatura sobre redes de conocimiento interorganizacionales se focaliza en organizaciones (por ejemplo, empresas, universidades, centros de investigación, entre otras) que se vinculan a través de lazos de conocimiento, ya sea para la transferencia, intercambio o cocreación de este. Los vínculos o lazos de conocimiento estudiados por la literatura son las alianzas estratégicas, los vínculos informales, la coautoría de publicaciones científicas y la coinvención de patentes, entre otros.

Las alianzas estratégicas (por ejemplo, acuerdos de I+D, licencias, consorcios de investigación o las transferencias de *know-how*) son vínculos de tipo formal –en los que media un contrato– muy estudiados para explorar las colaboraciones de conocimiento que involucran a empresas y/o organismos de ciencia y tecnología (es decir, acuerdos público-privados, privado-privado). En términos generales, las alianzas estratégicas pueden definirse como "arreglos voluntarios entre organizaciones que implican el intercambio, el uso común o el codesarrollo de productos, tecnologías y/o servicios" (Gulati, 1998). El crecimiento de este tipo de colaboraciones en la última década, así como la mayor accesibilidad de datos sobre estas, ha alentado el surgimiento de una literatura muy fértil sobre alianzas estratégicas dentro de los estudios de innovación y de negocios (Ozman, 2009; Meier, 2011).

La disponibilidad de datos también ha alentado el estudio de las colaboraciones interorganizacionales de conocimiento utilizando datos de coautoría de publicaciones científicas y de coinvención de patentes (Leydesdorff *et al*., 2012; Cantú *et al*., 2015; Capone, Innocenti y Lazzeretti, 2020). Sin embargo, cabe destacar, que este tipo de indicadores para aproximar lazos de conocimiento pueden no resultar del todo apropiados en casos de organizaciones involucradas en actividades poco propensas a publicar o patentar. Los vínculos de conocimiento de tipo informal, definidos como aquellos que implican compartir, recibir o crear conocimiento entre miembros de organizaciones sin que medie un contrato o registro formal de dicho lazo, son en general poco estudiados dadas las dificultades metodológicas que implica reconstruir este tipo de lazos. Sin embargo, estudios recientes han abordado las redes informales de conocimiento en distintas actividades (Conway, 1997; Fu Diez y Schiller, 2013; Stubrin, 2013; Ng y Law, 2015).

En cuanto al nivel de análisis, dos tipos de abordajes son los más frecuentes: el nivel egocéntrico y la red completa. En un análisis egocéntrico se analiza la red de vinculaciones de un nodo determinado. La unidad de análisis es el "ego de la red", el cual se puede describir por sus propios atributos y por sus vinculaciones. En una red de N actores, hay N unidades de análisis. En este tipo de abordajes, se decide teóricamente o empíricamente cuál es el ego, y luego se relevan los datos de vinculación de este (Arza *et al*., 2018). En el análisis de red completa se utiliza la información sobre los N actores que componen la red para analizarla de manera agregada. Desde un estudio "centrado en el ego", partiendo de ciertos nodos claves de la red, se puede ir reconstruyendo la red completa

a través del relevamiento de todas las *ego networks*. Cuando se trabaja con recolección de información primaria, usualmente utilizando una metodología de bola de nieve, se comienza por los nodos conocidos y se van generando nuevos nodos hasta llegar a la red completa.

Medición y datos

¿Cómo abordar metodológicamente un estudio sobre redes de conocimiento? En este apartado nos focalizamos principalmente en estilizar las diferentes estrategias que se utilizan en la literatura.

En primer lugar, en cuanto a las fuentes de datos, se utilizan tanto datos prefabricados o secundarios como autogenerados o primarios. La decisión metodológica de utilizar uno u otro está vinculada tanto a la disponibilidad como al objetivo de la investigación. Los datos prefabricados existentes permiten trazar redes de conocimiento a partir de lazos como coinvención de patentes, coautoría de publicaciones científicas y de alianzas estratégicas entre organizaciones. Las bases de datos disponibles sobre alianzas estratégicas en general no incluyen información sobre empresas en países en desarrollo, mientras que las bases de datos de patentes y publicaciones científicas no permiten una buena aproximación de las colaboraciones en las que participan organizaciones en sectores con baja propensión a publicar o patentar, o en países donde se patenta poco, como es el caso de países en desarrollo. Las bases de datos alternativas para utilizar, que empiezan a estar disponibles, son los registros de convenios de I+D, servicios técnicos y transferencias de conocimiento de universidades, por ejemplo, que pueden ser utilizadas para trazar un análisis egocéntrico de la red de conocimiento.

Ante la falta de datos prefabricados, las encuestas son los instrumentos más utilizados para generar los datos que permitan construir una red de conocimiento. Utilizamos el ejemplo de la red de conocimiento de empresas biotecnológicas en la Argentina (Stubrin, 2013) para ilustrar las principales decisiones metodológicas asociadas a un proceso de construcción de datos de redes. La inexistencia de un registro de alianzas estratégicas de este tipo de empresas en la Argentina llevó a la necesidad de planear un relevamiento a través de encuestas. En el caso estudiado, los nodos de la red fueron definidos utilizando la definición de empresa biotecnológica de la Organization

for Economic Co-operation and Development (OECD, 2005). A partir de una recopilación de bases de datos de empresas biotecnológicas existentes, otras fuentes secundarias y entrevistas, se obtuvo un primer padrón de empresas biotecnológicas en la Argentina, siendo un total de noventa empresas aproximadamente. La decisión fue la de estudiar la red completa. Es decir, estudiar las vinculaciones de todos los nodos de la red. Para asegurar que cada uno de los posibles nodos de la red cumpliera efectivamente con la definición de "empresa biotecnológica" adoptada, se procedió a incorporar en el formulario de la encuesta preguntas que permitieran confirmar que la empresa pertenecía a la red que se quería estudiar. Una estrategia de este tipo es necesaria cuando no se puede determinar *ex ante* la pertenencia de los nodos a la red que se quiere estudiar (por ejemplo, cuando los nodos son organizaciones que han participado de cierta actividad particular, o que tienen características difícilmente conocidas a través de bases de datos o información disponible).

En términos de relevamiento de vínculos, un método muy utilizado es el llamado "roster method" el cual consiste en presentarle a cada nodo entrevistado el listado de todos los otros posibles nodos y solicitarle que indique con cuál o cuáles ha tenido cierto tipo de vínculo (o qué frecuencia de vínculo ha tenido con cada uno de los otros nodos mencionados en el listado, por ejemplo). En la literatura, un estudio sobre redes e innovación que ha aplicado el "roster method" es el de Bell y Giuliani (2005) cuyo objetivo es entender la red de conocimiento de empresas vitivinícolas en el clúster del valle de Colchagua en Chile. Los nodos de la red fueron definidos como "empresas vitivinícolas que producen vino y la venden con el propio nombre de la bodega", identificándose veintiocho a través de registros de la región e informantes clave. Utilizando el "roster method" los autores realizaron una entrevista personal a los veintiocho nodos de la red en las que les mostraron un listado de los otros nodos de la red y les preguntaron por los vínculos de conocimiento con estos. Las preguntas utilizadas para identificar vínculos direccionados de asistencia técnica desde y hacia el nodo fueron: 1) "Si usted afronta una situación crítica y necesita de ayuda técnica, ¿a cuál de las firmas mencionadas en la lista acudiría? Identifique la importancia de la información obtenida por cada una en las siguientes categorías: 0 = ninguna, 1 = baja, 2 = media, 3 = alta" y 2) "¿Cuáles de las firmas en la lista usted cree que se han beneficiado de información técnica provista por su firma? Indique la importancia que

considera ha tenido la información transferida: 0 = ninguna, 1 = baja, 2 = media, 3 = alta". De esta manera, se obtuvieron datos de vínculos direccionados y ponderados de la red.

La técnica de recolección de datos descripta anteriormente no resulta adecuada cuando se cumplen algunas de las siguientes condiciones: se desconoce la población total de nodos de la red o la cantidad de nodos que componen la red es muy grande. En el primer caso, realizando un método de *roster* solo a un subconjunto conocido de la población, se corre el riesgo de obtener una red parcial que puede no incluir nodos que por sus atributos y/o vínculos son cruciales para comprender la estructura, dinámica y funcionamiento de la red. Por lo tanto, se corre el riesgo de llegar a una red que no refleja el fenómeno que queremos estudiar. En el segundo caso, cuando el N de la red es muy grande, proveer un listado de todos los nodos para que cada nodo marque aquellos con los que tuvo cierto vínculo e incluso brinde información sobre los vínculos (como frecuencia o importancia) con cada uno de los nodos puede ser tremendamente engorroso e ineficiente. Es por ello que en esos casos se utiliza un método alternativo de "generación de nodos" solicitando a cada nodo contactado que devele libremente (sin mostrarle un listado) quiénes son sus vínculos, y en algunos casos también los atributos de los vínculos y los nodos.

En el estudio de la red de conocimiento de empresas biotecnológicas en la Argentina (Stubrin, 2013) dado que se partía de un N muy alto (N = 90) y no se tenía certeza de que ese N reflejara la población total de nodos de la red, se optó por el método de "generación de nodos". Partiendo de los nodos conocidos, se le solicitó a cada uno de estos que revele los nombres de aquellas organizaciones con las que había tenido vínculos de conocimiento, de comercialización y producción. Los lazos de conocimiento se definieron como aquellos "contratos formales de I+D y licencias en que las empresas participaron en el período 2003-2008" y los flujos de comercialización/producción como "contratos formales de cooperación con fines de comercialización/producción en el período 2003-2008". Además de solicitar los nombres de las organizaciones con las cuales tuvo cierto tipo de vínculos, se pidió también que se revele la cantidad de vínculos que habían tenido con ese nodo en el período de tiempo del estudio.

Un aspecto de gran importancia en los relevamientos vía encuestas está vinculado a que la red que se termina obteniendo depende en gran medida de las respuestas al formulario enviado. En ese sentido, resulta

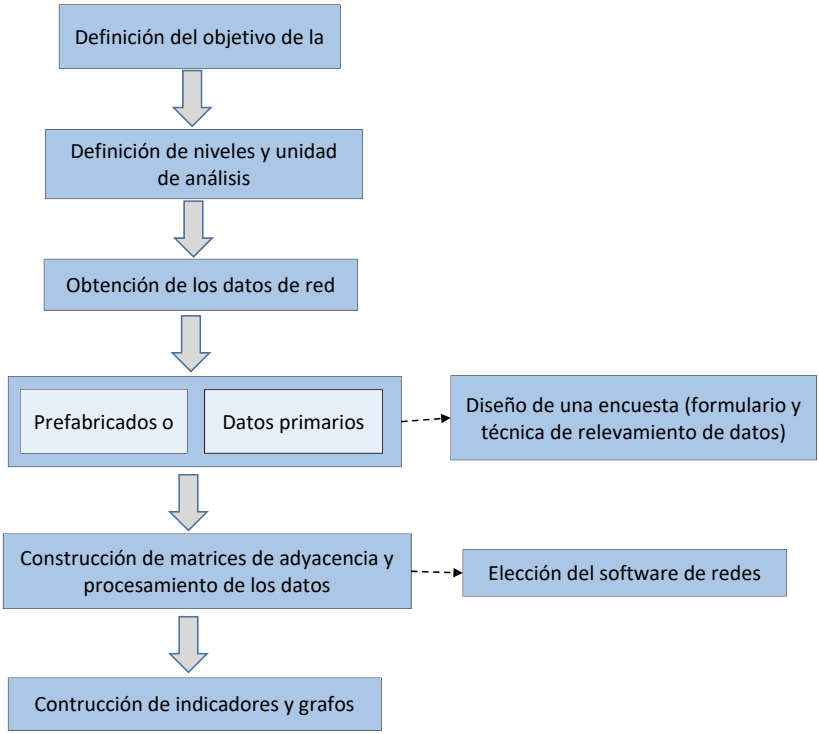
fundamental que las preguntas se realicen de manera tal que sean fácilmente interpretadas por quien responde la encuesta, y se deben realizar de manera tal de maximizar la capacidad de recordar y reconocer vínculos de quienes responden la encuesta. Un error frecuente es utilizar adjetivos como "importante" o "relevante" para consultar sobre los vínculos, lo que puede resultar en respuestas con alto nivel de subjetividad por parte de quien responde la encuesta ("¿con qué empresas de su sector intercambia sobre cuestiones relevantes a su trabajo?"). En algunas ocasiones ello puede aproximarse poniendo un tope a la cantidad de otros nodos que se le solicita a los nodos que se mencionen, asumiendo que estos seleccionarán los vínculos más estrechos ("por favor, mencione hasta cinco organizaciones que usted considere que hayan brindado conocimiento para resolver problemas técnicos en su organización en los últimos tres años").

Lograr el mayor grado de precisión en términos de las características del vínculo que se busca reconstruir y el período de tiempo que interesa abordar es fundamental para facilitar una correcta interpretación del vínculo por parte del entrevistado. Una posible estrategia para maximizar el recuerdo de todos los vínculos por parte de quienes responden la encuesta es preguntar de a un tipo de vínculo por vez y dejar en claro el tipo de nodo que se está buscando identificar (tipo de organización -empresa, universidad, laboratorio, etc.-, localización -en la misma región, en el país, en el extranjero, etc.-y otros atributos de interés). En muchas ocasiones también es posible a través de este método consultar sobre atributos de los nodos con quienes el ego colabora como nacionalidad, sector de pertenencia, entre otros. La posibilidad de administrar la encuesta de forma interpersonal (ya sea presencial o virtual) es recomendable dado que permite asegurarse de la correcta interpretación de los vínculos de la red por parte de quienes responden.

Descripción del método

En el diagrama 2 proveemos una descripción de la aplicación del ARS al estudio de las redes de conocimiento.

Diagrama 2. Descripción del ARS



Fuente: elaboración propia

Reflexiones finales

La complejidad de las nuevas tecnologías, el crecimiento de campos interdisciplinarios y el ritmo vertiginoso en que se produce el cambio técnico, entre otros, han alentado una creciente asociatividad y articulación entre distintos actores con el objetivo de intercambiar, transferir y crear conocimiento (Andrews, 1971; Brantley y Powell, 1992). Esto se manifiesta en parte en una expansión notoria en las últimas décadas de las colaboraciones científicas y de las redes de conocimiento interorganizacionales (Duysters y Hagedoorn, 2000; Wagner-Döbler, 2001; Fleming y Frenkel, 2007; Frenken, Ponds y van Oort, 2007; Acosta *et al*., 2011). En este capítulo hemos analizado cómo el ARS es un marco teórico y metodológico

que provee a los estudios de CTI de un herramental muy útil para entender mejor estas colaboraciones, y en particular cómo estas afectan la creación y difusión de conocimiento en América Latina y el Caribe (LAC).

Sin embargo, el aprovechamiento de la herramienta metodológica que brinda el ARS para el estudio de las colaboraciones de conocimiento en LAC enfrenta algunas limitaciones, las cuales están ligadas fundamentalmente a la disponibilidad de datos.

En el caso del estudio de las redes de colaboraciones científicas en LAC la principal limitación está vinculada al acceso a datos secundarios de calidad. El uso de banco bibliográficos internacionales disponibles tiende a subestimar el análisis de las redes de colaboraciones científicas en países de la región y su lugar en la producción de conocimiento a nivel mundial. En particular se señalan sesgos de idioma, de indexación de revistas locales, de acceso a los costos de publicación, entre otros. Si bien en la actualidad existen varios esfuerzos por mejorar la indexación de la producción de artículos en la región a partir de repositorios como SciELO, Latindex o LA Referencia, estos bancos aún no permiten la extracción detallada de metadatos para realizar ARS.

Varios autores han señalado la importancia de trabajar con otras fuentes de información para la región que permitan mayor foco en las capacidades nacionales de producción de conocimiento, un ejemplo es el uso de plataformas nacionales de currículos para el análisis de redes de colaboración a partir de datos bibliográficos (Digiampietri *et al*., 2014), proyectos de investigación (Bianchi *et al*., 2019) y tutorías académicas en tesis de posgrado (Mena-Chalco y Rossi, 2014). Si bien estas fuentes de datos presentan limitantes en torno al tiempo de procesamiento y homogeneización, así como el carácter autodeclarado de los datos, su uso tiene el potencial de captar formas de colaboración más diversa y con foco en las particularidades de los sistemas de CTI de los países latinoamericanos.

En referencia al estudio de las redes de conocimiento interorganizacionales en LAC, el acceso a datos es aún más difícil. Las bases de datos sobre alianzas estratégicas que son pocas a nivel mundial tienen un acceso mucho más restringido y poseen una baja o nula representatividad de organizaciones situadas en LAC. Las bases de datos más difundidas –Cooperative Agreements and Technology Indicators (CATI), Alliances, Recombinant Capital (RECAP), BioScan o Advanced Research Project on Agreement– se nutren de la recopilación de la publicación de anuncios públicos de acuerdos de colaboración en periódicos y revistas de

negocios especializadas.[8] En general, la cobertura está sesgada a grandes empresas de países europeos, asiáticos o norteamericanos. Por lo tanto, el conocimiento de las redes de conocimiento interoganizacionales en la región, a través de vínculos como alianzas estratégicas, demanda la construcción de datos primarios a través de encuestas. El alto costo de este tipo de investigaciones limita la producción de estudios sobre el tema. Ello repercute, por lo tanto, en una literatura aún muy incipiente que nos da todavía un bajo nivel de conocimiento acerca de la estructura y funcionamiento de las redes de conocimiento interorganizacionales en LAC, y del impacto que estas tienen sobre la capacidad de innovación de empresas, sectores y países.

Concluimos este capítulo afirmando que en la aplicación del ARS al estudio de las redes de conocimiento en LAC hay un enorme espacio para la contribución de nuevas investigaciones, y esperamos que este capítulo contribuya a la difusión y mayor aplicación de esta metodología en los estudios de CTI en la región.

Bibliografía

Acosta, M.; Coronado, D.; Ferrándiz, E. y León, M. D. (2011). "Factors affecting inter-regional academic scientific collaboration within Europe: The role of economic distance". *Scientometrics*, vol. 87, nº 1, pp. 63-74.

Adshead, M. y Quayle, M. (2018). "The resilience of regional African HIV/AIDS research networks to the withdrawal of international authors in the subfield of public administration and governance: Lessons for funders and collaborators". *Scientometrics*, vol. 117, nº 1, pp. 163-173.

Ahuja, G. (2000). "Collaboration networks, structural holes, and innovation: A longitudinal study". *Administrative science quarterly*, vol. 45, nº 3, pp. 425-455.

8 La base de datos CATI es producida por el instituto de investigación holandés UNU-MERIT, mientras que Alliances pertenece a la agencia Thompson Reuters. Estas bases de datos cubren alianzas estratégicas en todos los sectores. Otras como Recombinant Capital (RECAP), BioScan y Advanced Research Project on Agreement son sector-específicas. Las dos primeras en biotecnología, mientras que la tercera en alianzas en el área de TIC.

Ainsworth, S.; Cortés, H. D.; del Río, J. A.; Narváez-Berthelemot, N. y Russell, J. M. (2007). "Colaboración científica entre países de la región latinoamericana". *Revista Española de Documentación Científica*, vol. 30, nº 2, pp. 180-198.

Albert, R. y Barabási, A. L. (1999). "Emergence of Scaling in Random Networks". *Science*, vol. 286, nº 5439, pp. 509-512.

Andrews, K. (1971). *The Concept of Corporate Strategy*. Homewood, IL: Irwin.

Araújo, T. y Fontainha, E. (2017). "The specific shapes of gender imbalance in scientific authorships: a network approach". *Journal of Informetrics*, vol. 11, nº 1, pp. 88-102.

Argote, J. G.; Collazo, G. E. E. y Rivero, A. A. G. (2017). "Producción científica sobre enfermedades infecciosas desatendidas en Latinoamérica". *Revista Electrónica Dr. Zoilo E. Marinello Vidaurreta*, vol. 42, nº 5.

Arza, V.; López, E.; Marín, A. y Stubrin, L. (2018). "Redes de conocimiento asociadas a la producción de recursos naturales en América Latina: análisis comparativo". *Revista CEPAL*, nº 125.

Barabási, A.-L. (2002). *Linked: The New Science of Networks.* New York: Perseus Books.

Batagelj, V.; De Nooy, W. y Mrvar, A. (2018). *Exploratory social network analysis with Pajek: Revised and expanded edition for updated software*. Cambridge: Cambridge University Press.

Baum, J. A.; Rowley, T. J. y Shipilov, A. V. (2003). "Where do small worlds come from?". *Industrial and Corporate change*, vol. 12, nº 4, pp. 697-725.

Beaver, D. (2001). "Reflections on scientific collaboration (and its study): past, present, and future". *Scientometrics*, vol. 52, nº 3, pp. 365-377.

_____ (2013). "The many faces of collaboration and teamwork in scientific research: updated reflections on scientific collaboration". *COLLNET Journal of Scientometrics and Information Management*, vol. 7, nº1, pp. 45-54.

Bell, M. y Giugiliani, E. (2005). "The micro-determinants of meso-level learning and innovation: evidence from a Chilean wine cluster". *Research policy*, vol. 34, nº 1, pp. 47-68.

Bianchi, C.; Couto Soares, M. C. y Tomassini Urti, C. (2019). "Health-related knowledge production in Brazil: Regional interaction networks and priority setting". *Innovation and Development*, vol. 9, nº 2, pp. 187-204.

Bonacich, P. (1972). "Factoring and weighting approaches to status scores and clique identification". *Journal of mathematical sociology*, vol. 2, nº 1, pp. 113-120.

Borgatti, S. P.; Brass, D. J.; Labianca, G. y Mehra, A. (2009). "Network analysis in the social sciences". *Science*, vol. 323, nº 5916, pp. 892-895.

Borgatti, S. P.; Everett, M. G. y Johnson, J. C. (2013). *Analyzing Social Networks*. Thousand Oaks, CA: SAGE.

Borgatti, S. P.; Perry, B. L. y Pescosolido, B. A. (2018). *Egocentric network analysis: Foundations, methods, and models*, vol. 44. Cambridge: Cambridge University Press.

Brantley, P. y Powell, W. W. (1992). "Networks and organizations: structure, form, and action". *Competitive Cooperation in Biotechnology: learning through networks*.

Burt, R. S. (1995). *Structural holes*. Cambridge, MA: Harvard University Press.

_____ (2004). "Structural holes and good ideas". *American journal of sociology*, vol. 110, nº 2, pp. 349-399.

Cantú, C.; Corsaro, D.; Dagnino, G. B.; Levanti, G.; Minà, A.; Picone, P. M. y Tunisini, A. (2015). "Interorganizational network and innovation: a bibliometric study and proposed research agenda". *The journal of business & industrial marketing*, vol. 30, nº 3-4, pp. 354-377.

Capone, F.; Innocenti, N. y Lazzeretti, L. (2020). "Knowledge networks and industrial structure for regional innovation: An analysis of patents collaborations in Italy". *Papers in Regional Science*, vol. 99, nº 1, pp. 55-72.

Chavarro, D.; Rafols, I. y Tang, P. (2017). "To What Extent is Inclusion in the Web of Science an Indicator of Journal 'quality'?". *Research Evaluation*, vol. 27, nº 2, pp. 106-118.

Cherven, K. (2015). *Mastering Gephi network visualization*. Birmingham, Reino Unido: Packt.

Cohanoff, C.; Mena-Chalco, J. P.; Robaina, S. y Tomassini, C. (2021). "Health Research Networks Based on National CV Platforms in Brazil and Uruguay". *Journal of Scientometric Research*, vol. 10, nº 1s.

Coleman, J. S. (1988). "Social capital in the creation of human capital". *American journal of sociology*, vol. 94, pp. S95-S120.

Consoli, D. y Mina, A. (2009). "An evolutionary perspective on health innovation systems". *Journal of Evolutionary Economics*, vol. 19, nº 2, pp. 297-319.

Consoli, D. y Ramlogan, R. (2008). "Out of sight: problem sequences and epistemic boundaries of medical know-how on glaucoma". *Journal of Evolutionary Economics*, 18(1), 31-56.

_____ (2012). "Patterns of organization in the development of medical know-how: the case of glaucoma research". *Industrial and corporate change*, vol. 21, nº 2, pp. 315-343.

Conway, S. (1997). "Informal networks of relationships in successful small firm innovation". *Technology, Innovation and Enterprise*, pp. 236-273. Londres: Palgrave Macmillan.

Costa, B. M. G.; da Silva Pedro, E. y de Macedo, G. R. (2013). "Scientific collaboration in biotechnology: the case of the northeast region in Brazil". *Scientometrics*, vol. 95, nº 2, pp. 571-592.

Cowan, R. (2004). *Network models of innovation and knowledge diffusion*. USA: MERIT.

Cowan, R. y Jonard, N. (2004). "Network structure and the diffusion of knowledge". *Journal of Economic Dynamics and Control*, vol. 28, nº 8, pp. 1557-1575.

_____ (2007). "Structural holes, innovation and the distribution of ideas". *Journal of Economic Interaction and Coordination*, vol. 2, nº 2, pp. 93-110.

_____ (2008). *If the Alliance Fits...: Innovation and Network Dynamics.* Países Bajos: UNU-MERIT.

Cronin, B.; La Barre, K. y Shaw, D. (2003). "A cast of thousands: Coauthorship and subauthorship collaboration in the 20th century as manifested in the scholarly journal literature of psychology and philosophy". *Journal of the American Society for information Science and Technology*, vol. 54, nº 9, pp. 855-871.

de Paula Fonseca e Fonseca, B.; Fernandes, E. y Fonseca, M. V. A. (2017). "Collaboration in science and technology organizations of the public sector: A network perspective". *Science and Public Policy*, vol. 44, nº 1, pp. 37-49.

de Paula Fonseca e Fonseca, B.; Guindalini, C.; Lopes Fonseca, F.; Machado-Silva, A. y Pereira-Silva, M. V. (2019). "Scientific and technological contributions of Latin America and Caribbean countries to the Zika virus outbreak". *BMC Public Health*, vol. 19, nº 1.

Digiampietri, L. A.; Martins Lopes, F.; Mena-Chalco, J. P. y Roberto Marcondes, C. (2014). "Brazilian bibliometric coauthorship networks". *Journal of the Association for Information Science and Technology*, vol. 65, nº 7, pp. 1424-1445.

dos Reis Azevedo Botelho, M.; Ruffoni, J.; Stefani, R. y Tatsch, A. L. (2022). "Knowledge networks in Brazil's health sciences". *Science and Public Policy*, vol. 49, nº 1, pp. 72-84.

Duque, R. B.; Dzorgbo, D. B. S.; Mbatia, P.; Shrum, W.; Sooryamoorthy, R. e Ynalvez, M. (2005). "Collaboration paradox: Scientific productivity, the Internet, and problems of research in developing areas". *Social studies of science*, vol. 35, nº 5, pp. 755-785.

Durrance, J. C. y Williams, K. (2008). "Social networks and social capital: Rethinking theory in community informatics". *The Journal of Community Informatics*, vol. 4, nº 3.

Duysters, G. y Hagedoorn, J. (2000). "Organizational modes of strategic technology partnering". *Journal of Scientific & Industrial Research*, vol. 59, nº 8-9, pp. 640-649.

Duysters, G. y Verspagen, B. (2004). "The small worlds of strategic technology alliances". *Technovation*, vol. 24, nº 7, pp. 563-571.

Dyer, J. H. y Nobeoka, K. (2000). "Creating and managing a high-performance knowledge-sharing network: the Toyota case". *Strategic management journal*, vol. 21, nº 3, pp. 345-367.

Eisenhardt, K. M. y Schoonhoven, C. B. (1996). "Resource-based view of strategic alliance formation: Strategic and social effects in entrepreneurial firms". O*rganization Science*, vol. 7, nº 2, pp. 136-150.

Erdős, P. y Rényi, A. (1961). "On the Strength of Connectedness of a Random Graph". *Acta Mathematica Academiae Scientiarum Hungaricae*, vol. 12, pp. 261-267. DOI: https://doi.org/10.1007/BF02066689.

Faust, K. y Wasserman, S. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Fleming, L. y Frenken, K. (2007). "The evolution of inventor networks in the Silicon Valley and Boston regions". *Advances in Complex Systems*, vol. 10, n° 1, pp. 53-71.

Freeman, L. C. (1979). "Centrality in Social Networks Conceptual Clarification". *Social Networks*, vol. 1, pp. 215-239.

_____ (2012). *El desarrollo del análisis de redes sociales.: un estudio de sociología de la ciencia*. Bloomington: Palibrio.

Frenken, K.; Ponds, R. y van Oort, F. (2007). "The geographical and institutional proximity of research collaboration". *Papers in regional science*, vol. 86, n° 3, pp. 423-443.

Fu, W., Diez, J. R., y Schiller, D. (2013). "Interactive learning, informal networks and innovation: Evidence from electronics firm survey in the Pearl River Delta, China". *Research Policy*, vol. 42, n° 3, pp. 635-646.

Gargiulo, M. y Gulati, R. (1999). "Where do interorganizational networks come from?". *American journal of sociology*, vol. 104, n° 5, pp. 1439-1493.

Gibbons, M.; Limoges, C.; Nowotny, H.; Schwartzman, S.; Scott, P. y Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. Thousand Oaks, CA: SAGE.

Gibbons, M.; Nowotny, H. y Scott, P. (2001). *Re-Thinking Science: Knowledge and the Public in an Age of Uncertainty*. Cambridge, Reino Unido: Polity.

González-Brambila, C. N.; Krackhardt, D. y Veloso, F. M. (2013). "The impact of network embeddedness on research output". *Research Policy*, vol. 42, n° 9, pp. 1555-1567.

Granovetter, M. (1973). "The strength of weak ties". *American journal of sociology*, vol. 78, n° 6, pp. 1360-1380.

_____ (1985). "Economic action and social structure: The problem of embeddedness". *American journal of sociology*, vol. 91, nº 3, pp. 481-510.

Gulati, R. (1998). "Alliances and networks". *Strategic management journal*, vol. 19, nº 4, pp. 293-317.

Hagedoorn, J. y Schakenraad, J. (1992). "Leading companies and networks of strategic alliances in information technologies". *Research policy*, vol. 21, nº 2, pp. 163-190.

Hicks, D. (2004). "The four literatures of social science". En Glänzel, W.; Moed, H. F. y Schmoch, U. (eds.), *Handbook of Quantitative Science and Technology Research*, pp. 473-496. Dordrecht: Springer.

Hidalgo, C. A. (2016). "Disconnected, fragmented, or united? A trans-disciplinary review of network science". *Applied Network Science*, vol. 1, nº 1, pp. 1-19.

Jackson, M. O. (2010). *Social and economic networks*. Princeton: Princeton University Press.

Kastelle, T. y Steen, J. (2010). "Are small world networks always best for innovation?". *Innovation*, vol. 12, nº 1, pp. 75-87.

Katz, J. S. (1994). "Geographical proximity and scientific collaboration". *Scientometrics*, vol. 31, nº 1, pp. 31-43.

Katz, J. S. y Martin, B. R. (1997). "What is research collaboration?". *Research Policy*, vol. 26, nº 1, pp. 1-18.

Kong, X.; Liu, J.; Shi, Y.; Xia, F. y Yu, S. (2019). "Academic social networks: Modeling, analysis, mining and applications". *Journal of Network and Computer Applications*, vol. 132, pp. 86-103.

Koput, K. W.; Owen-Smith, J.; Powell, W. W. y White, D. R. (2005). "Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences". *American journal of sociology*, vol. 110, nº 4, pp. 1132-1205.

Koput, K. W.; Powell, W. W. y Smith-Doerr, L. (1996). "Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology". *Administrative science quarterly*, pp. 116-145.

_____ (1997). "Strategies of learning and industry structure: the evolution of networks in biotechnology". *Advances in Strategic Management*, vol. 14, pp. 229-254.

Koskinen, J.; Lusher, D. y Robins, G. (eds.) (2012). *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications*. Cambridge: Cambridge University Press

Laudel, G. (2001). "Collaboration, creativity and rewards: Why and how scientists collaborate". *International Journal of Technology Management*, vol. 22, n° 7-8, pp. 762-781.

_____ (2002). "What do we measure by co-authorships?". *Research evaluation*, vol. 11, n° 1, pp. 3-15.

Lazarsfeld, P. F. y Merton, R. K. (1954). "Friendship as a social process: A substantive and methodological analysis". *Freedom and control in modern society*, vol. 18, n° 1, pp. 18-66.

Leta, J.; Sorenson, M. M. y Vasconcelos, S. M. R. (2009). "A new input indicator for the assessment of science & technology research?". *Scientometrics*, vol. 80, n° 1, pp. 217-230.

Leydesdorff, L. y Rafols, I. (2009). "A Global Map of Science Based on the ISI Subject Categories". *Journal of the American Society for Information Science and Technology*, vol. 60, n° 2, pp. 348-362. DOI: https://doi.org/10.48550/arXiv.0911.1057.

Leydesdorff, L. y Wagner, C. S. (2005). "Network structure, self-organization, and the growth of international collaboration in science". *Research Policy*, vol. 34, n° 10, pp. 1608-1618.

Leydesdorff, L.; Nightingale, P.; O'Hare, A.; Rafols, I. y Stirling, A. (2012). "How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business & management". *Research policy*, vol. 41, n° 7, pp. 1262-1282.

Leydesdorff, L.; Porter, A. L. y Rafols, I. (2010). "Science overlay maps: A new tool for research policy and library management". *Journal of the American Society for Information Science and Technology*, vol. 61, n° 9, pp. 1871-1887. DOI: https://doi.org/10.1002/asi.21368.

Li, E. Y.; Liao, C. H. y Yen, H. R. (2013). "Co-authorship networks and research impact: A social capital perspective". *Research Policy*, vol. 42, n° 9, pp. 1515-1530.

Luke, D. A. (2015). *A user's guide to network analysis in R*. New York: Springer.

Meier, M. (2011). "Knowledge management in strategic alliances: A review of empirical evidence". *International journal of management reviews*, vol. 13, nº 1, pp. 1-23.

Mena-Chalco, J. P. y Rossi, L. (2014). "Caracterização de árvores de genealogia acadêmica por meio de métricas em grafos". *Anais do III Brazilian Workshop on Social Network Analysis and Mining*. Porto Alegre: Sociedade Brasileira de Computação.

Molina González, J. L. (2005). "El estudio de las redes personales: contribuciones, métodos y perspectivas". *Empiria. Revista de metodología de ciencias sociales*, nº 10, pp. 71-106.

Moreno, J. L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama*. Reino Unido: Beacon House

Mowery, D. C.; Oxley, J. E. y Silverman, B. S. (1998). "Technological overlap and interfirm cooperation: implications for the resource-based view of the firm". *Research policy*, vol. 27, nº 5, pp. 507-523.

Newman, M. E. J. (2001). "The structure of scientific collaboration networks". *Proceedings of the National Academy of Sciences*, vol. 98, nº 2, pp. 404-409.

_____ (2006). "Modularity and community structure in networks". *Proceedings of the National Academy of Sciences*, vol. 103, nº 23, pp. 8577-8582.

Ng, D. C. W. y Law, K. (2015). "Impacts of informal networks on innovation performance: evidence in Shanghai". *Chinese Management Studies*, vol. 9, nº 1, pp. 56-72. DOI: https://doi.org/10.1108/CMS-05-2013-0077.

Organization for Economic Co-operation and Development (OECD) (2005). "Chapter 2: Basic Concepts and Definitions". *A Framework for Biotechnology Statistics*. París: OECD Publishing.

Ozman, M. (2009). "Inter-firm networks and innovation: a survey of literature". *Economic of Innovation and New Technology*, vol. 18, nº 1, pp. 39-67.

Pammolli, F. y Riccaboni, M. (2002). "On firm growth in networks". *Research Policy*, vol. 31, nº 8-9, pp. 1405-1416.

Pérez Beltrán, J. E.; Rodríguez-Aceves, L. y Valerio Ureña, G. (2015). "Análisis de redes sociales para el estudio de la producción intelectual

en grupos de investigación". *Perfiles educativos*, vol. 37, n° 150, pp. 124-142.

Strogatz, S. H. y Watts, D. J. (1998). "Collective dynamics of 'small-world' networks". *Nature*, vol. 393, n° 6684, pp. 440-442.

Stubrin, L. I. (2013). *High-Tech activities in emerging countries: a network perspectives of the Argentinean biotech activity*. Maastricht: Universitaire Pers Maastricht.

Wagner-Döbler, R. (2001). "Continuity and discontinuity of collaboration behaviour since 1800 from a bibliometric point of view". *Scientometrics*, vol. 52, n° 3, pp. 503-517.

White, H. (1992). *Identity and Control: A Structural Theory of Social Action*. Princeton: Princeton University Press.

Ziman, J. (2000). *Real Science: What it is and what it means.* Cambridge: Cambridge University Press.

Capítulo 3
Aplicaciones de la teoría de grafos al análisis de sistemas de innovación y espacios tecnológicos

Ana Urraca Ruiz, Pedro Miranda y Vanessa de Lima Avanci

Introducción

Hay dos características esenciales en la naturaleza del progreso técnico. La primera es su carácter evolutivo, esto es, el avance del conocimiento, independientemente de cuál sea su ámbito, se caracteriza por estar en permanente transformación. La segunda es que el progreso técnico es el resultado del funcionamiento de un sistema altamente complejo en el que interaccionan dos espacios o subsistemas: el de innovación-producción y el tecnológico.

Cuando nos referimos al progreso técnico como resultado de interacciones entre y dentro de los sistemas, una forma de aproximarnos a sus características y evolución es observarlo en términos de una red, esto es, de objetos y de interacciones entre objetos. De esta forma, los sistemas de innovación-producción y tecnológico son diferentes en la medida en que constituyen redes de interacciones entre diferentes objetos. El espacio de innovación-producción está constituido por las interrelaciones entre los objetos que participan del proceso de innovación-difusión, concretamente, el *agente innovador* y el *resto de los agentes*, esto es, empresas, universidades, centros públicos de investigación, fundaciones y organizaciones no gubernamentales, administraciones públicas y familias. La literatura llama este espacio como "sistemas de innovación" en dimensión territorial –nacional, regional o local– y sectorial. El segundo espacio está constituido

por las interacciones entre *piezas de conocimiento* para un estado dado del avance científico y tecnológico. Las piezas de conocimiento pueden estar referidas a un paradigma científico (por ejemplo, microbiología, física cuántica, química orgánica, etc.); a una tecnología específica o artefacto (por ejemplo, el Boeing 747, un tratamiento para el sida o un programa que permite transmitir datos, etc.); a un conjunto de tecnologías que se encuentran dentro del mismo paradigma tecnológico (por ejemplo, aerodinámica, antibióticos, tecnologías digitales, etc.) o a un conjunto de tecnologías que combinan conocimiento de diferentes paradigmas (por ejemplo, tecnologías de transporte, tecnologías ambientales, etc.).

La teoría de grafos representa una metodología de análisis apropiada para estudiar fenómenos que suceden por interacción tomando la forma de un sistema en red. Un grafo es una estructura discreta que representa una red de interrelaciones (jerárquicas o no jerárquicas) entre objetos. De acuerdo con el tipo de interacción entre los objetos, los grafos pueden ser de dos tipos: dirigidos y no dirigidos. En los grafos dirigidos, los objetos están relacionados siguiendo una dirección o jerarquía, mientras que en los no dirigidos no hay jerarquía en la relación entre objetos. En principio, los grafos no dirigidos tienen una aplicabilidad mayor en sistemas de innovación en los que no existe jerarquía en las relaciones entre objetos, como son los casos de la cooperación tecnológica o de las bases de conocimiento. Alternativamente, los grafos dirigidos son más aplicables a fenómenos en los que hay acumulatividad y direccionalidad, como, por ejemplo, trayectorias tecnológicas o flujos de conocimiento desde un origen a un destino, como es el caso de *spillovers*, transferencia y captura tecnológica entre agentes o países.

Las fuentes de información que permiten la representación de un sistema mediante un grafo son de diferente naturaleza. Para representar sistemas de innovación se pueden utilizar datos procedentes de encuestas específicas locales, bases de datos oficiales de carácter catastral relativos a grupos de investigación, bases de datos de artículos científicos, depósitos de patentes en conjunto o las matrices insumo-producto. Las bases de conocimiento se representan, casi exclusivamente, con datos de patentes (depósitos y citas) (Jaffe y Trajtenberg, 2002; Maurseth y Verspagen; 2002; Breschi, Lissoni y Malerba, 2003;

Kushnir, Leydesdorff y Rafols, 2014; Kogler, Leydesdorff y Yan, 2017), aunque también podría utilizarse la bibliometría (publicaciones de artículos científicos). En los países latinoamericanos existe una propensión a patentar relativamente más baja, lo que convierte los indicadores de

patentes en un indicador poco representativo de la capacidad de innovación de estos países. Sin embargo, para realizar prospectiva tecnológica en determinadas trayectorias tecnológicas para tecnologías específicas pueden ser utilizados, pues para este tipo de estudio se utilizarían datos de patentes del mundo. Para estudios más específicos de países concretos, a pesar de tener un bajo número de patentes, la metodología de grafos puede ser utilizada para mapear el origen del conocimiento y cualificar algunos aspectos de dependencia tecnológica.

Este capítulo tiene como objetivo presentar algunas nociones básicas de teoría de grafos y cómo se pueden construir grafos para representar y analizar bases de conocimiento. Para esto, el capítulo se organiza en tres apartados además de esta introducción. En el segundo apartado se presenta de forma muy general qué es un grafo, sus tipos y sus propiedades. Luego revisamos las principales fuentes de información que pueden ser utilizadas para construir grafos, haciendo un especial énfasis en las patentes por ser estas la principal fuente utilizada para construir bases de conocimiento tecnológico. El último apartado muestra cómo construir grafos de bases de conocimiento a partir de las informaciones de las patentes y algunos indicadores para analizar sus propiedades. Finalmente, se presentan las conclusiones del capítulo con algunas reflexiones y recomendaciones finales.

Grafos y análisis de redes

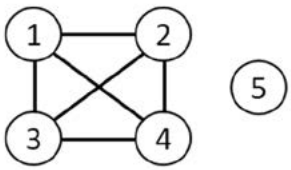
Un grafo G representa una red de interrelaciones entre un conjunto no vacío de nodos (o vértices, V) a través de un conjunto de aristas (o arcos, E) tal que $G = (V,E)$. En un grafo *simple*, cada arista (e-E) conecta un par no ordenado de nodos diferentes $\{x, y\}$ tal que $\{x, y\} - VxV$. Como el par no es ordenado, la representación $\{x, y\}$ es la misma que $\{y, x\}$ y se dice que el grafo es no dirigido.

Los nodos son los objetos que forman el grafo y cada uno lleva asociado un *grado* (g) determinado por el número de aristas que lo tienen como extremo. El número de nodos que contiene un grafo representa su *orden*, esto es, $G = |V|$. Un grafo es regular si todos sus nodos tienen el mismo grado. El número de aristas que contiene un grafo representa su *tamaño* (figura 1). Las aristas pueden ser de tres tipos: *ramas*, *paralelas* y *lazos*. Las ramas son las aristas comunes que unen dos nodos. Las paralelas son aristas conjuntas cuyos nodos inicial y final son el mismo. Finalmente,

los lazos, también llamados aristas cíclicas o bucles, son aristas que relacionan el nodo con él mismo. Si en un grafo hay aristas paralelas, se llama multigrafo, y, si además hay lazos, se llama pseudografo. Si cada par de nodos del grafo está conectado por una arista y además el grafo es regular, esto es, todos sus nodos tienen grado n -1, entonces el grafo es completo. Un grafo completo con n nodos tendrá siempre $n(n$ -1)/2 aristas. Por otro lado, si un grafo simple tiene nodos no conectados, se denomina grafo no completo.

Los grafos tienen cuatro propiedades. La primera es la "adyacencia". Dos aristas son adyacentes cuando tienen un nodo en común y dos nodos son adyacentes o vecinos cuando están unidos por una arista. La segunda es la "incidencia". Una arista es incidente a los nodos $\{x, y\}$ cuando los une. La tercera es la "ponderación", la cual sucede cuando hacemos corresponder una función que asocia a cada arista un valor o peso (w) tal que $w(e)$-R, el cual le permite aumentar su expresividad en la red. En este caso, el grafo pasa a ser ponderado: $G = (V,E, w)$. La cuarta propiedad es el "etiquetado" que permite distinguir nodos o aristas específicos mediante algún rasgo que los haga distinguibles del resto.

Figura 1. Ejemplo de grafo simple no completo: orden, tamaño y grado (g)



$V = \{1,2,3,4,5\}$

$E = \{\{1,2\},\{1,3\},\{1,4\},\{2,3\},\{2,4\},\{3,4\}\}$

$Orden: 5;\ Tamaño: 6$

$g(1) = 3;\ g(2) = 3;\ g(3) = 3;\ g(4) = 2;\ g(5 = 0)$

Fuente: elaboración propia

Aunque existe una amplia variedad y complejidad de grafos,[1] este capítulo se va a centrar en los dos tipos con mayor aplicabilidad al estudio de bases de conocimiento: pseudografos dirigidos y no dirigidos. Un pseudografo es dirigido cuando la adyacencia entre dos nodos es unívoca, esto

1 Para un estudio en profundidad de los tipos de grafos, árboles, representación y aplicaciones, recomendamos la lectura de manuales especializados como son los de Grimaldi (1997), Biggs (1998), Aho, Hopcroft y Ullman (1998), Rosen (1999), Bogart (2000), Grassmann y Tremblay (2003) o Caicedo, Mendez y Wagner de García (2010).

es, existe un nodo cola y un nodo cabeza. Si la arista que los conecta es de tipo nodo inicial –nodo terminal, el nodo cabeza es adyacente al nodo cola. En este caso, la red expresa una direccionalidad en la relación entre los objetos. En los grafos no dirigidos, las aristas expresan relaciones biunívocas entre nodos. En este caso, dos nodos son adyacentes si están conectados por una arista y cada uno de ellos es, a su vez, incidente con ella. Pasamos a examinar las propiedades de cada uno de ellos.

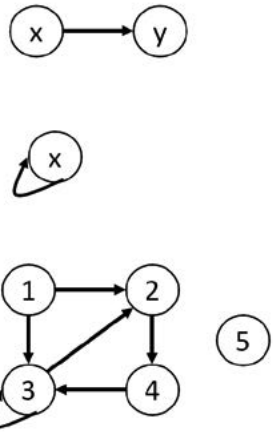
Grafos dirigidos

También llamados dígrafos. En este tipo de grafos las aristas pueden denominarse también arcos o líneas dirigidos. En los grafos dirigidos las aristas representan un par ordenado de vértices {*x*, *y*} en el que *x* es el nodo inicial e *y* el nodo terminal. En el primer ejemplo de la figura 2, la arista va de *x* a *y* (*x* –*y*) y el nodo *y* es adyacente e incidente al nodo *x*. El segundo ejemplo de la figura 2 es un lazo o bucle, en el que la arista une el nodo *x* con él mismo. En este caso la arista va de *x* a *x* (*x* –*x*) y el nodo *x* es adyacente e incidente a él mismo.

En los grafos dirigidos se diferencia, para cada nodo, entre el grado de entrada $g^-(v)$ y el grado de salida $g^+(v)$. El grado de entrada de *v* es el número de aristas que tienen a *v* como nodo terminal y el grado de salida es el número de aristas que tienen a *v* como un nodo inicial (véase el tercer ejemplo de la figura 2).

Los grafos dirigidos definen caminos o cadenas, esto es, secuencias entre los nodos $v_1, v_2, ..., v_n$, definidas por las aristas que los conectan y del tipo v_1–v_2, v_2–v_3, ..., v_{n-1}–v_n. La longitud del camino es el número de aristas que hay en el camino, es decir, $n - 1$ para un camino linear con *n* nodos. Sin embargo, en un camino se pueden repetir los nodos y las aristas. Cada nodo define un camino de longitud 0 entre él y él mismo (si no hubiera lazos). Un camino es *simple* si todas las aristas son diferentes (aunque se repitan nodos); un *circuito* es un camino simple que empieza y termina en el mismo nodo; un *camino elemental* tiene todos los nodos diferentes excepto el primero o el último; un ciclo es un camino elemental que empieza y termina en el mismo nodo.

Figura 2. Grafos dirigidos



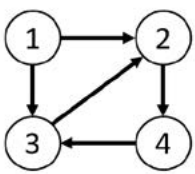
$x \to y$

$x \to x$

$g^-(1) = 0 \quad g^+(1) = 2$
$g^-(2) = 2 \quad g^+(2) = 1$
$g^-(3) = 3 \quad g^+(3) = 2$
$g^-(4) = 1 \quad g^+(4) = 1$
$g^-(5) = 0 \quad g^+(5) = 0$

Fuente: elaboración propia

Figura 3. Caminos en grafos dirigidos



$V = \{1,2,3,4\}$

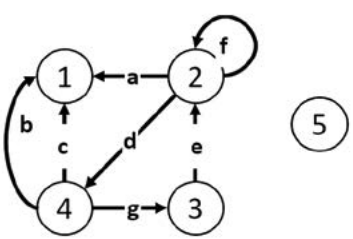
$E = \{\{1,2\},\{1,3\},\{3,2\},\{2,4\},\{4,3\}\}$

camino: (1,3,2,4,3,2); *circuito*: (3,2,4,3)

camino elemental: (3,2,4); *ciclo*: (3,2,4,3)

Fuente: elaboración propia

Figura 4. Ejemplo de matriz de adyacencia de un pseudografo dirigido



$A_G =$

	1	2	3	4	5
1	0	0	0	0	0
2	1	1	0	1	0
3	0	1	0	0	0

Fuente: elaboración propia

La forma de operar matemáticamente un grafo es mediante su representación matricial en las llamadas "matrices de adyacencia". La matriz de adyacencia de un grafo $G(V,E)$, denotada por (A_G), tiene dimensión $n \times n$, siendo n el número de nodos del grafo. Cada celda (a_{ij}) representa el número de aristas que van del nodo i al nodo j. Las matrices de adyacencia no son necesariamente simétricas. Si G es un grafo simple, $a_{ii} = 0$ y $a_{ij} - [0,1\}$ y si es un multigrafo, $a_{ii} = 0$.

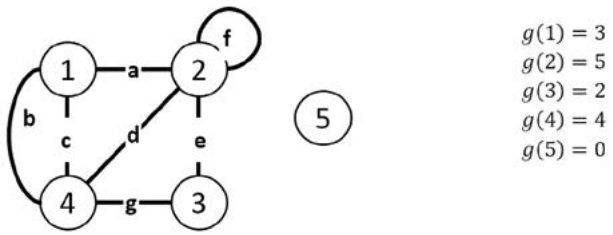
Grafos no dirigidos

El grado de un nodo en un grafo no dirigido es el número de aristas incidentes que contiene. Los lazos dan dos unidades al grado del nodo que lo contiene. Los nodos de grado 1 se denominan terminales y los de grado 0, aislados (véase figura 5).

Los grafos no dirigidos también definen caminos en sus interrelaciones. En este caso, el camino se define como la secuencia no ordenada de nodos que van del nodo 1 hasta el nodo n, pudiéndose repetir nodos y aristas. Los caminos pueden ser igualmente simples, elementales, circuitos o ciclos, siendo que no se consideran ciclos aquellos caminos de longitud cero (x); de longitud 1, es decir, cuando un nodo tiene un lazo (v,v); o de longitud 2, es decir, cuando dos nodos definen aristas paralelas (x, y,x).

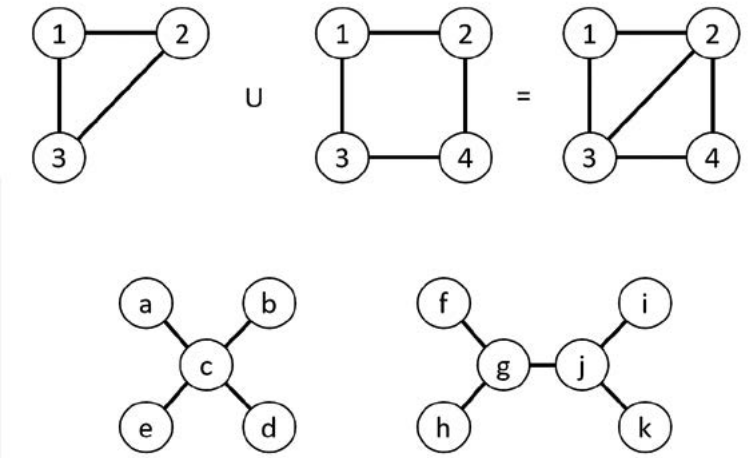
Un grafo no dirigido puede dividirse en subgrafos. Se dice que $G^- = (V^-,E^-)$ es un subgrafo de G si V^- es un subconjunto de V y si E^- contiene las aristas $\{x, y\}$ en E y en V^-. La unión de dos subgrafos simples $G_1^- = (V_1^-,E_1^-)$ y $G_2^- = (V_2^-,E_2^-)$ es un grafo simple $G = G_1^- \cup G_2^- = (V_1^- \cup V_2^-, E_1^- \cup E_2^-)$ (figura 6). Los grafos no dirigidos pueden ser *conexos*, si todos sus pares de vértices están conectados por algún camino elemental, y *no conexos*, cuando están formados por la unión de varios subgrafos conexos y desconectados entre sí, en cuyo caso se llaman *componentes conexos* del grafo (figura 6).

Figura 5. Grado en pseudografos no dirigidos



Fuente: elaboración propia

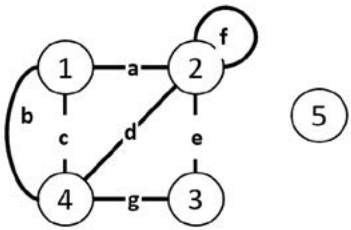
Figura 6. Unión de subgrafos simples y grafos no conexos



Fuente: elaboración propia

Existen cuatro formas de representar grafos no dirigidos: matriz de adyacencia, definida en los mismos términos que en los grafos dirigidos; matriz de adyacencia etiquetada; matriz de incidencia y lista de adyacencia. La matriz de adyacencia tiene dimensión $n \times n$ y cada componente a_{ij} cuenta el número de aristas que une los nodos $i - j$. Como en estos grafos las artistas no tienen dirección, la matriz de adyacencia es simétrica, es decir, las relaciones $i - j$ deben ser iguales a las relaciones $j - i$ (figura 7). Si G es un grafo simple, los valores de la diagonal (a_{ii}) deben ser iguales a 0 y para el resto de los valores a_{ij} –{0,1}. Si es un multigrafo, apenas los valores de la diagonal serán iguales a 0. La matriz de adyacencia etiquetada es aquella en la que las celdas a_{ij} son ponderaciones de la arista que une los nodos $i - j$. La matriz de incidencia del grafo $G(V,E)$, denotada como I_G, es una matriz de orden $|V| \times |E|$, en la que cada celda I_{ij} toma valor 0 si la arista j no es incidente al nodo i y valor 1 si lo es. Finalmente, un grafo no dirigido se puede representar como uno dirigido sustituyendo la arista no direccionada $\{x, y\}$ en dos aristas direccionadas, una de x a y y otra de y a x.

Figura 7. Ejemplo de matriz de adyacencia de un grafo no dirigido

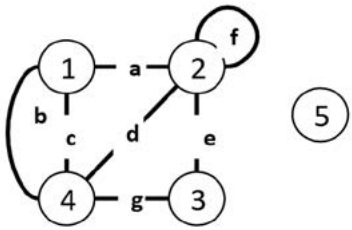


$A_G =$

	1	2	3	4	5
1	0	1	0	2	0
2	1	1	1	1	0
3	0	1	0	1	0

Fuente: elaboración propia

Figura 8. Ejemplo de matriz de incidencia de un grafo no dirigido



$I_G =$

	a	b	c	d	e	f	g
1	1	1	1	0	0	0	0
2	1	0	0	1	1	1	0
3	0	0	0	0	1	0	1

Fuente: elaboración propia

Análisis de redes: centralidad, conectividad y segmentación

Además de las medidas presentadas, como el tamaño y el grado, otras estadísticas ayudan a describir la estructura de las redes (grafos) y sus elementos (nodos). Dichas medidas son especialmente relevantes en redes grandes y elevada densidad, cuya visualización es compleja y su interpretación difícil. Estas medidas sirven además para comparar redes y clasificarlas según las propiedades observadas.

Las principales medidas de caracterización de la red son las siguientes (Jackson, 2010; Carrington y Scott, 2011; Barabási, 2016):

1. Distribución de grado: esta es una característica fundamental de una red que describe la frecuencia relativa de los nodos que tienen diferentes grados. Entonces *P(d)* indica la fracción de nodos en la red que tienen grado *d*.
2. Densidad: la densidad representa el nivel de interconexión entre los nodos de una red y puede ser calculada como la relación entre la cantidad de enlaces entre pares de nodos y el total de enlaces posibles. Se define como:

$$D = \frac{2l}{n \times (n-1)}$$

3. Diámetro y longitud media del camino: la distancia entre dos nodos (*i*, *j*) de una red viene dada por el menor número de enlaces entre ambos (*w*). Cuando dos nodos (*i*, *j*) están directamente conectados, la distancia entre ellos es 1, [$w(i,j) = 1$] (el camino más corto posible). Sin embargo, cuando dos nodos no están directamente relacionados, la distancia entre ellos es la suma de los enlaces directos que intermedian esta relación. Si no hay camino entre los nodos, la distancia entre ellos es infinita. Después de calcular el camino más corto entre todos los pares de nodos en una red, el diámetro es la distancia más larga entre ellos. Por lo tanto, el diámetro de una red, *D**, se define como la distancia más grande entre dos nodos, tal que:

$$D^* = \max_{i \neq j} w(i,j)$$

   La medida de distancia caracteriza pares de nodos. En términos agregados, la distancia media del camino indica la media del menor camino entre todos los diferentes pares de nodos que componen la red.

4. Coeficiente de agrupamiento medio de red:

- El coeficiente de agrupamiento medio permite clasificar y comparar el grado de cohesión entre los nodos. La distribución del coeficiente de agrupamiento individual puede variar entre redes. El coeficiente de agrupamiento de un vértice C_i viene dado por la

relación entre el número de bordes entre los nodos adyacentes y el número máximo posible de bordes entre nodos adyacentes. Supongamos, por ejemplo, los enlaces ij e ik. En este caso, los nodos *j* y *k* son adyacentes al nodo *i*. La medida de agrupamiento para *i* viene dada por la proporción de veces en que la relación jk ocurre, dado el número de veces en que ocurren las relaciones ij e ik. El coeficiente de agrupamiento medio se calcula como el promedio de los coeficientes de agrupamiento individual C_i de todos los nodos de la red (*n*), tal que: $\overline{C} = \sum_{i = 1}^{n}\frac{C_{i}}{n}$

- Además de las medidas presentadas anteriormente, hay un conjunto de medidas de centralidad que representa las propiedades de los nodos. Estas medidas indican cómo cada uno se relaciona con los demás e identifican los más centrales en la estructura de la red (Freeman, 1978; Strogatz y Watts, 1998; Borgatti, 2005). En este grupo están la centralidad de grado, de proximidad y de intermediación. Otras medidas de representatividad de los nodos se derivan de la centralidad de vector propio como el cálculo del PageRank de Google, utilizado para medir la relevancia de páginas web en Internet (Bonacich, 1972)

5. Centralidad de grado: el grado es una medida que muestra el nivel de interconexión de un nodo en la red. Un nodo es central si está conectado a muchos otros nodos. En una red de tamaño *n*, el grado máximo posible de cualquier nodo *i* es igual a (*n* -1). Por lo tanto, la centralidad de grado de un nodo *i* (G_i^*) estará dada por la relación entre su grado ($g(i)$) y su valor máximo:

$$G_i^* = \frac{1}{n-1} \times g(i)$$

G_i^* varía, toma valor 0 cuando *i* es un nodo aislado y 1 cuando está conectado con todos los demás nodos de la red.

6. Centralidad de proximidad: la centralidad de proximidad también indica cuán conectado está un nodo a la red, pero tiene en cuenta no solo las conexiones directas, sino también las indirectas, esto es, las intermediadas por otros nodos. Los nodos más importantes de la red, según este criterio, son aquellos que tienen las distancias

más cortas de otros nodos. La medida se basa en la distancia promedio entre un nodo dado y todos los demás nodos en la red. En una red con n nodos, la centralidad de proximidad del nodo i se puede calcular como:

$$P_i = \frac{(n-1)}{\sum_{j \neq i} w(i,j)}$$

En la que $w(i, j)$ es la distancia más pequeña entre los nodos i y j.

7. Centralidad de intermediación: los nodos con mayor valor de intermediación en una red son los que actúan como puentes acortando el camino entre otros nodos. El indicador de intermediación para i se define de la siguiente manera:

$$B_i = \sum \frac{w_{j,k}(i)}{w_{j,k}} ; k \neq j \neq i$$

La centralidad de intermediación (B_i) se calcula como la razón entre la cantidad de veces que i estuvo en la ruta más corta entre los demás pares de nodos $w_{j,k}(i)$ y la cantidad de caminos más cortos entre j y k ($w_{j,k}$). Un valor de B_i próximo a 1 significa que i está presente en casi todas las rutas más cortas que conectan j y k. Lo contrario ocurre cuando B_i se acerca a 0.

Fuentes de información

Para crear el grafo que represente una base de conocimiento se necesita información que identifique los objetos (nodos) y sus relaciones (aristas). Existen diversas fuentes en las que se puede encontrar este tipo de información (cuadro 1). Algunas fuentes ya presentan la información en formato matricial y pueden ser adaptadas para ser tratadas como matrices de adyacencia direccionadas o no direccionadas. En otras ocasiones, la información deberá ser tratada previamente antes de ser colocada en formato matricial. Y en otras, será necesario consultar diversas fuentes para recolectar las informaciones necesarias para montar el grafo.

Cuadro 1. Principales fuentes de información

Sistema	Fuente
Sistemas de innovación	Fuentes primarias: Encuestas específicas de carácter sectorial o territorial Fuentes secundarias: • Composición de grupos de investigación (agencias nacionales de investigación) • ORBIS, Who Owns Whom (identificación de vínculos y naturaleza de la propiedad) • Publicaciones especializadas sobre acuerdos de cooperación, joint ventures, etc. • Websites de las propias empresas • Patentes y modelos de utilidades (depósitos en conjunto entre empresas sin vínculo de propiedad) • Tablas *input-output* (sistemas de cuentas nacionales) • Estadísticas de empleo, educación, etc.
Sistemas tecnológicos o bases de conocimiento	• Depósitos de patentes (oficinas nacionales de patentes,* EPO y PATSTAT) • Citas de patentes (oficinas nacionales de patentes, EPO y PATSTAT) • Publicaciones científicas (Web of Science, Scopus) • Citaciones de publicaciones científicas (Web of Science, Scopus)

Fuente: elaboración propia

Nota: (*) Para obtener información sobre las oficinas nacionales de patentes, se puede consultar la página web de la World Intellectual Property Organization (WIPO) (https://www.wipo.int/members/en/).

Los grafos que representan redes relativas a sistemas de innovación se pueden construir con fuentes de información primarias o secundarias. Las fuentes primarias pueden incluir preguntas en los cuestionarios relativas a interacciones con otros agentes de forma que esta información pueda ser traducida a una expresión matricial. Entre las fuentes de carácter secundario, destacamos aquellas que reportan indirectamente una red, como, por ejemplo, las tablas *input-output* para el análisis de relaciones intersectoriales. En otros casos, las matrices se construyen con información cruzada. Por ejemplo, si quisiéramos analizar cómo se incorpora el progreso técnico al salario, podríamos construir matrices de incidencia en las que el tipo de cualificación profesional es común a un par de sectores. Los registros de grupos de investigación en agencias de fomento nacionales, de autores de artículos científicos, o los depósitos de patentes también se pueden utilizar para estudiar fenómenos como la cooperación tecnológica o la transferencia de tecnología. En estos casos, es posible que se necesiten fuentes adicionales para identificar perfectamente los nodos. Por ejemplo, en el caso de la cooperación, para confirmar

la inexistencia de vínculo de propiedad entre los agentes envueltos en la cooperación, sería necesario utilizar fuentes adicionales como ORBIS, la base Who Owns Whom o los websites de las propias empresas. Estas mismas bases sirven también para identificar la principal clasificación industrial del depositante, dado que esta es una información no provista por las bases de datos de patentes.

Las principales fuentes de información utilizadas para construir grafos relativos a bases de conocimiento deben contener alguna información que permita identificar las piezas de conocimiento, siendo este un concepto altamente abstracto y de difícil medición. En términos empíricos, la forma de identificar piezas de conocimiento es a través de clasificaciones que diferencian campos o disciplinas científico-tecnológicas para algún nivel de agregación. En este sentido, las bases de patentes y de publicaciones científicas permiten identificar dominios de conocimiento a través de la clasificación internacional de patentes y de palabras clave. Las citas de patentes, así como las citas de artículos científicos, permiten elaborar grafos direccionados sobre formación de tecnologías o avances científicos a partir de conocimiento previo o por nuevas combinaciones entre piezas de conocimiento. También permiten generar grafos que identifiquen la direccionalidad e intensidad de flujos tecnológicos interregionales (por ejemplo, conocimiento generado en un país a partir del generado en otros). Sin embargo, las estadísticas de patentes no recogen todas las características de una base de conocimiento en sentido estricto e introducen algunos fallos de información. Pasamos a analizar algunos de ellos.

Patentes y bases de conocimiento

Una patente representa una tecnología en la forma de un nuevo producto, proceso o método. Independientemente de cuál sea la oficina en que esté depositada, cada patente tiene asignado uno o más dominios tecnológicos que representan funcionalidades interconectadas. Estos dominios tecnológicos están representados por códigos de acuerdo con la clasificación internacional de patentes (IPC, por sus siglas en inglés)[1]

1 Los códigos IPC representan dominios tecnológicos, pues se asignan a las patentes de acuerdo con las características tecnológicas o la aplicación del objeto patentado. La IPC no es el único sistema de clasificación, pero es el adoptado por más de cien oficinas de patentes en todo el mundo.

que se refieren a funcionalidades (áreas de conocimiento tecnológico) vinculadas a disciplinas de conocimiento científico (por ejemplo, química, farmacia, mecánica, etc.). Cuanto mayor sea el nivel de especificidad de la subtecnología, mayor será el nivel de desagregación y más específico será el campo de aplicación (cuadro 2).[2]

Los códigos IPC están organizados en un sistema jerárquico, alfanumérico seguido de pontos, en los que el título de cada subclase describe las características principales de la parte del conjunto de conocimientos (cuadro 2).[3] Considerar los códigos IPC para un cierto nivel de desagregación como unidades de conocimiento tecnológico implica una cierta arbitrariedad. Dado que el conocimiento es un sistema complejo, su divisibilidad en partes más pequeñas combinadas es un artificio, ya que, para mayores niveles de desagregación, las subunidades se subdividen y combinan de forma igualmente compleja.

2 La IPC tiene una estructura jerárquica de quince niveles. Los primeros niveles se identifican como sección (1 dígito), clase (1 a 3 dígitos), subclase (1 a 4 dígitos), grupo y subgrupo. Cada subgrupo, a su vez, puede estar compuesto por hasta once subniveles jerárquicos, que se identifican por puntos. Por lo tanto, cada símbolo de la IPC puede tener hasta catorce dígitos, seguidos de hasta once puntos. Para obtener más información sobre la IPC, consultar en WIPO: https://www.wipo.int/classifications/ipc/en/.

3 Los títulos de cada símbolo deben leerse teniendo en cuenta toda la estructura jerárquica. De esta manera, el título del símbolo H01F 1/053 será "imanes de materiales inorgánicos caracterizados por su coercitividad, que comprende aleaciones magnéticas duras, que contienen específicamente metales de tierras raras" (WIPO, 2019).

Cuadro 2. Ejemplo de la IPC y de las funcionalidades tecnológicas para diferentes niveles de agregación

H	Electricidad
H01	Elementos eléctricos básicos
H01B	Cables; conductores; aisladores; empleo de materiales específicos por sus propiedades conductoras, aislantes o dieléctricas...
H01C	Resistencias
H01F	Imanes; inductancias; transformadores; empleo de materiales específicos por sus propiedades magnéticas
H01F 1/00	Imanes o cuerpos magnéticos (caracterizados por los materiales magnéticos pertinentes); empleo de materiales específicos por sus propiedades magnéticas
H01F 1/01	• de materiales inorgánicos...
H01F 1/03	•• caracterizados por su coercitividad H01F 1/032 ••• de materiales magnéticos duros H01F 1/04 •••• metales o aleaciones
H01F 1/047	••••• aleaciones caracterizadas por su composición
H01F 1/053	•••••• que contienen metales de tierras raras

Fuente: elaboración propia

Fuente: Oficina Española de Patentes y Marcas. Disponible en: http://pubcip.oepm.es/classifications/ipc/ipcpub/?notion=scheme&version=20200101&symbol=none&menulang=es&lang=es&viewmode=f&fipcpc=no&showd

La principal ventaja de usar bases de patentes para representar bases de conocimiento es que los códigos IPC pueden agregarse a partir de características de conocimiento comunes. Además, los datos de patentes permiten análisis para largos períodos de tiempo y para un conjunto amplio de países sin necesidad de restricciones a campos técnicos o sectores específicos (ver fuentes mencionadas en el cuadro 1). Sin embargo, también existen ciertas limitaciones. En primer lugar, no todos los esfuerzos de innovación resultan en patentes, ya sea porque no toda tecnología es patentable, porque la expectativa de retorno económico es baja en comparación con los costes de patentar, porque las empresas usan otros mecanismos de apropiación (secreto industrial, ciclo de vida reducido del producto, marketing, etc.) o porque los sistemas de patentes no son sistemas de apropiación efectivos, lo que sucede en industrias específicas.

Además, la normativa sobre derechos de propiedad industrial e intelectual es diferente entre países, por lo que la propensión a patentar varía entre regiones. En el caso de los países de América Latina, además de la debilidad institucional de sus sistemas de apropiabilidad, existe una baja propensión a innovar por parte de los agentes nacionales, y una menor aun propensión a patentar. Para la concesión de una patente, se deben cumplir tres requisitos básicos: novedad, actividad inventiva y aplicación industrial. En países alejados de la frontera tecnológica, parte del esfuerzo innovador está asociado con la difusión o adaptación de tecnologías ya existentes, por lo que no se genera un objeto patentable.[4] Así, el número de depósitos de países latinoamericanos es bajo en relación con los países tecnológicamente más avanzados, aunque similar a países de nivel tecnológico medio como los de Europa del Este (Milesi, Petelski y Verre 2017) en los casos de Brasil, México, la Argentina o Chile. En este sentido, las bases de conocimiento tecnológico nacionales representadas apenas por patentes tenderán a estar subrepresentadas. Este problema es aún más relevante cuando se utilizan bases de datos de otros países, como la United States Patent and Trademark Office (USPTO) o la European Patent Office (EPO), dado que la presentación de patentes en países extranjeros está influenciada por las relaciones comerciales y el grado de internacionalización de las empresas. Existen algunas soluciones para ampliar el número de observaciones. Dado que el criterio de selección de patentes para la base de conocimiento nacional es la residencia del inventor y que las oficinas nacionales tienen un cierto "sesgo doméstico", es preferible utilizar datos de oficinas nacionales o regionales o con familias de patentes, las cuales incluyen todas las patentes con inventor residente en el país de estudio independientemente de la oficina en que fue depositada.[5] En este caso, es necesario llevar en consideración que los países poseen legislaciones y procedimiento específicos. Otra forma de ampliar la base de conocimiento nacional en América Latina sería estudiar conjuntamente una agregación de grafos relativos a patentes, modelos de utilidad y publicaciones.

En segundo lugar, las patentes también están sujetas a un sesgo de distribución sectorial debido a las diferentes propensiones a patentar por industria y tecnología. Lo mismo ocurre cuando se considera el tamaño

4 Para indicadores basados en estadísticas de patentes en general, ver el *OECD-Patent statistics manual* (2009).

5 En este caso, es necesario tener en cuenta que los países tienen diferentes leyes y procedimientos.

de las empresas, dado que los costes de los procedimientos de patentamiento y de eventuales disputas legales para mantener los derechos de la patente pueden ser excesivos para las empresas más pequeñas.

En tercer lugar, los códigos IPC representan dominios de aplicación del conocimiento, no piezas de conocimiento en sentido estricto. En este sentido, son una representación limitada de lo que serían "piezas de conocimiento". En teoría, una base de conocimiento está construida por piezas que mantienen sus propios atributos de conocimiento. Sin embargo, cuando las bases de patentes representan bases de conocimiento, se pierden algunos de estos atributos. Por ejemplo, el conocimiento puede ser tácito o codificado. Sin embargo, las patentes solo representan conocimiento codificado y solo si este tiene alguna aplicación industrial o funcionalidad. El conocimiento tácito no puede ser clasificado y el conocimiento científico no se ajusta a la clasificación IPC, por lo que para representar conocimiento científico se deben utilizar grupos de palabras clave adecuados. El conocimiento es también interdisciplinar en función de la similitud y complementariedad entre piezas de conocimiento. En las representaciones con bases de patentes, la interdisciplinariedad adquiere una dimensión diferente; esta sería la propiedad que adquiere un campo técnico (código IPC) o una cierta funcionalidad de combinarse con otros campos técnicos que pertenecen a diferentes dominios tecnológicos. En este sentido, la interdisciplinariedad tiene un significado de ubicuidad, proximidad entre diferentes funcionalidades o convergencia entre subtecnologías para crear otra tecnología.

En cuarto lugar, debe considerarse que las patentes no tienen el mismo valor económico ni representan el mismo grado de avance, esto es, mientras que algunas representan innovaciones radicales, otras son aplicaciones o innovaciones incrementales. Además, algunas patentes puede que nunca lleguen a las líneas de producción. Esto se debe a que las empresas las utilizan solo para establecer "vallas de patentes" alrededor de sus productos principales. Con este tipo de estrategia, las empresas pueden restringir la actuación tecnológica de sus competidores en áreas tecnológicas próximas o afines a las propias; pueden evitar disputas legales o usar las patentes como un activo en la negociación de licencias cruzadas con competidores; o pueden utilizarlas como un indicador de rendimiento de sus departamentos de I+D para comercializar sus productos o para atraer inversores. A pesar de esta heterogeneidad en la naturaleza de las patentes, todas se contabilizan por igual, esto es, tendrán el mismo peso para representar bases de conocimiento.

En quinto lugar, los códigos de clasificación generalmente son asignados por examinadores expertos que deben identificar el área de conocimiento de la tecnología representada por la patente en la amplitud de los dominios de conocimiento. Sin embargo, la interpretación de las especificaciones técnicas para asignar una o varias clasificaciones a la patente puede variar entre examinadores.

Finalmente, debido a que la IPC es una clasificación aplicada *ex post*, solo incorpora nuevas unidades de conocimiento cuando es revisada, lo cual sucede con una cierta temporalidad y de forma artificial. La periodicidad de la revisión de la IPC puede atrasar la visualización de nuevos nodos en la red y cambios en su estructura (Kay *et al*., 2014). Una forma de contornear este problema es utilizar buscadores lexicográficos de palabras clave en la descripción de la patente. Estas palabras clave serían nuevos dominios o tecnologías disruptivas que se agregarían como nuevos vértices.

La base de conocimiento creada a partir de una base completa de patentes representaría una base agregada que recoge todo el conocimiento acumulado desde su año de inicio. Esta base agregada podría desglosarse según el objeto de análisis en bases de conocimiento industrial (de industrias específicas según la actividad principal del depositante), nacional (de países específicos según el lugar de residencia del inventor), temporal (períodos específicos de acuerdo con la fecha de presentación) y tecnológica (tecnologías específicas o grupos tecnológicos para ciertas agregaciones tecnológicas como FISIR (OST, 2010), tecnologías ambientales, KET (de Heide *et al*., 2015), etc.).

Aplicaciones de la teoría de grafos al análisis de bases de conocimiento

Un espacio tecnológico o base de conocimiento es una estructura organizada (no aleatoria) e interrelacionada de objetos, subunidades o piezas de conocimiento tecnológico (Krafft *et al*., 2011). La forma como esas piezas de conocimiento se estructuran revela una *funcionalidad*, esto es, la respuesta a problemas concretos de acuerdo con el "estado del arte" de métodos, procesos, habilidades y técnicas disponibles en un momento dado de tiempo. Por ejemplo, podemos hablar de base de conocimiento de las tecnologías de "vehículos eléctricos híbridos", que es una posible tecnología de "vehículos eléctricos", la cual se encuentra dentro de la base de "tecnologías de emisión

de bajo carbono" que, a su vez, pertenecen a la base de conocimiento de las tecnologías ambientales y tecnologías de transporte.

Las bases de conocimiento son, por tanto, redes altamente complejas y pueden ser agregadas para organizaciones específicas (empresa, universidad, *joint ventures*, conglomerados, etc.), industrias, períodos y territorios (países, regiones, etc.). Las bases de conocimiento referidas a industrias representan agregaciones de las competencias centrales de las firmas que operan en ellas (productivas, tecnológicas y organizativas) y sus activos complementares. En este sentido, el grado de diversificación de la base de conocimiento industrial estará fuertemente relacionado con los caminos de diversificación de las empresas que componen el sector. En la medida en que las bases de conocimiento industrial representan competencias y capacidades, ellas nos permiten estudiar fenómenos como las trayectorias tecnológicas industriales (corte temporal), el grado de permeabilidad o ubiquidad de las tecnologías (*pervasiveness*), los procesos de convergencia tecnológica entre industrias, las motivaciones para la cooperación interindustrial por disimilitud y complementariedad entre competencias tecnológicas, etc.

Las bases de conocimiento también pueden estar referidas a territorios agrupando las bases de conocimiento referentes a las organizaciones que residen en ellos. Las bases de conocimiento de orden territorial permiten estudiar la especialización tecnológica de un espacio económico o político respecto de una zona de referencia, sus ventajas tecnológicas y su proceso de formación de competencias a lo largo del tiempo. A su vez, se puede estudiar la aportación de cada uno de los agentes del sistema a la base de conocimiento espacial desagregando la base nacional en subgrafos relativos a la base de conocimiento de empresas (nacionales y multinacionales), inventores independientes, universidades y centros públicos de investigación, fundaciones y organizaciones no gubernamentales y agencias o departamentos de gobierno.

Finalmente, las bases de conocimiento de orden temporal, de forma agregada o para algún nivel de desagregación (tecnología, industria, región, etc.) permiten observar cómo se transforma el grafo a medida que el progreso técnico avanza, bien por discontinuidad, es decir, por el surgimiento de nuevos paradigmas o microparadigmas, o bien por continuidad, esto es, por la recombinación de las piezas de conocimiento que integran la red (Krafft *et al*., 2011; Nesta y Saviotti, 2005; Saviotti, 2009). El estudio de la transformación de la red sirve para realizar prospección tecnológica, puesto que tecnologías muy próximas en un

espacio tecnológico es la representación de un elevado potencial de recombinación para formar una trayectoria (tecnología) o un camino de diversificación tecnológica (región, industria o empresa) (Kajikawa *et al*., 2015). En este sentido, este tipo de estudios puede servir de auxilio para la articulación de políticas de innovación, mediante la identificación de la falta de conectividad, *trade-offs*, necesidad de convergencia entre dominios tecnológicos, convivencia y competencia entre microparadigmas, etc., así como la forma como estos fenómenos coevoluyen.

Representación de bases de conocimiento como grafos

El conocimiento es un sistema complejo, es decir, las piezas que lo componen no son perfectamente indivisibles, sino que representan combinaciones de otras subpiezas de conocimiento para algún nivel de desagregación. Eso significa que dominios tecnológicos específicos pueden ser representaciones concretas de las piezas de conocimiento que componen una estructura de conocimiento en abstracto. Para realizar una equivalencia entre las piezas o unidades de conocimiento tecnológico (en el plano abstracto) y los dominios tecnológicos observados (en el plano concreto), se deben formular dos supuestos. El primero es asumir que los dominios son las piezas de conocimiento que lo componen y comparten semejanzas de acuerdo con algún criterio de clasificación tecnológica y para algún grado de agregación. Las piezas de conocimiento vendrían representadas por los códigos IPC para algún grado de desagregación, si se utilizan datos de patentes (cuadro 3). El segundo supuesto supone que cuando una tecnología combina diferentes dominios tecnológicos define relaciones entre diferentes tipos de conocimiento. Estas relaciones representan las conexiones del grafo (aristas). De esta forma, podemos construir un grafo que representa una base de conocimiento tecnológico en la cual: 1) la incorporación de nuevos dominios tecnológicos (nuevos nodos) significa que la base está agregando nuevas piezas de conocimiento; 2) el tamaño de un nodo será mayor cuanto mayor sea el uso del dominio tecnológico asociado; 3) la recombinación de dominios tecnológicos lleva al surgimiento de nuevas aristas y 4) la densidad del grafo en términos de número de conexiones (aristas) aumenta cuanto mayor sea la frecuencia con la que se combinan dos dominios tecnológicos.

Para construir el grafo asociado a una base de conocimiento utilizando datos de patentes se debe escoger, en primer lugar, el nivel de desagregación

de los códigos IPC que representan los nodos. El número de nodos determina las dimensiones de la matriz de adyacencia e irá aumentando con el nivel de desagregación. La IPC, en su versión de 2020, presenta ocho secciones (1 dígito), 131 clases (3 dígitos), 646 subclases (4 dígitos), 7518 grupos principales (10 dígitos) y 68.030 subgrupos (10-14 dígitos). La elección por el nivel de desagregación con el que se va a trabajar dependerá del objeto de análisis. Cuando se investigan tecnologías específicas para hacer prospección tecnológica se suele trabajar con niveles de desagregación elevados, mientras que cuando se estudian bases de conocimiento de agentes, industrias o regiones se suele trabajar con niveles de agregación de 3-4 dígitos.

Cuadro 3. Elementos teóricos de la base de conocimiento y metodología de análisis empírico

Elementos teóricos	**Teoría de grafos**	**Fuente de información**
Nivel abstracto	*Instrumento*	*Nivel concreto*
Base de conocimiento	Grafo	Bases de patentes agregadas
Piezas de conocimiento	Nodos	*Dimensiones de la matriz de adyacencia* • códigos IPC (2-14 dígitos) • palabras clave (lexicografía) • códigos de otras clasificaciones de acuerdo con la funcionalidad
Relaciones de conocimiento	Aristas	*Celdas de la matriz de adyacencia* • diversas medidas de distancia entre clases tecnológicas diferentes • lazos y nodos inconexos

Fuente: elaboración propia

Una vez determinadas las dimensiones de la matriz de adyacencia, es necesario definir un método apropiado para representar las aristas o enlaces entre nodos. Existen numerosas formas de establecer enlaces entre los nodos, aunque la literatura no aclara cuál sería la mejor. Esta ambigüedad es un límite al uso de grafos para realizar prospección tecnológica, pues la medida utilizada determina la estructura del grafo. El trabajo de Luo y Yan (2016) realiza una revisión de la literatura sobre cómo asignar los enlaces y el grado de sus nodos asociados a partir de diferentes medidas. Los autores establecen cuatro categorías de acuerdo con el método y fuente de información utilizados: 1) por citas de patentes; 2) por la probabilidad de que un mismo inventor (organizaciones, individuos o países) deposite

una patente en un par de campos técnicos; 3) por la frecuencia con que un agente patenta en un par de campos técnicos y 4) por la frecuencia con que dos campos técnicos coocurren en la misma patente. Para cada una de estas categorías, los autores definen diversas medidas de enlace (ver cuadro 4).

Cuadro 4. Estimación de las aristas como medidas ponderadas de distancia entre nodos

Categoría	Fuente	Medida	Definición	Aplicaciones
(1) Semejanza (relatedness)	Citas de patentes	1. Cocita normalizada	Número de citaciones compartidas normalizadas por el número de todas las citas únicas de patentes en un par de clases.	Mapeamiento de redes de patentes
		2. Semejanza del coseno clase a clase	El coseno del ángulo de los dos vectores que representan dos distribuciones de citas de clases tecnológicas en todas las clases de patentes.	Redes tecnológicas
		3. Semejanza del coseno clase a patente	El coseno del ángulo de los dos vectores que representan dos distribuciones de citas de clases tecnológicas en patentes únicas.	Mapeamiento de redes de patentes (mejorado)
(2) Probabilidad de un agente inventor en diversos campos de conocimiento	Bibliometría o información bibliográfica del inventor asignada por región	1. Probabilidad de diversificar del inventor	Menor probabilidad condicional (entre el par) de que un inventor esté especializado en patentar en una clase cuando también está especializado en la otra.	Inicialmente usados para diversificación de regiones (espacios-producto). Aplicabilidad a la diversificación tecnológica y redes tecnológicas
		2. Probabilidad de diversificar de una organización	Menor probabilidad condicional (entre el par) de que una organización esté especializada en patentar en una clase cuando está especializada en la otra.	
		3. Probabilidad de diversificar de un país/región	Menor probabilidad condicional (entre el par) de que un país/región esté especializado en patentar en una clase cuando también está especializado en la otra.	
(3) Frecuencia de la diversificación de un agente entre campos técnicos	Bibliometría o información bibliográfica del inventor asignada por región	1. Frecuencia de la coocurrencia del inventor	La desviación entre el número de inventores asociados a un par de clases y el valor esperado suponiendo que los patrones de diversificación son aleatorios.	Inicialmente usados para relatedness industrial. Aplicación para medir la frecuencia de innovación de los agentes por coocurrencia
		2. Frecuencia de la coocurrencia de una organización	La desviación entre el número de organizaciones asociadas a un par de clases y el valor esperado suponiendo que los patrones de diversificación son aleatorios.	
		3. Frecuencia de la coocurrencia de un país/región	La desviación entre el número de países inventores asociados a un par de clases y el valor esperado suponiendo que los patrones de diversificación son aleatorios.	
(4) Frecuencia de campos técnicos que comparten la misma patente	Información de múltiples clases asignadas a una patente	1. Coclasificación o coocurrencia normalizada	Número de patentes compartidas de un par de clases tecnológicas normalizadas por el número de patentes únicas en ambas clases.	Mapeamiento de redes de patentes
		2. Similaridad del coseno de la coclasificación o coocurrencia	El coseno del ángulo de los dos vectores que representan dos distribuciones de dos clases tecnológicas de patentes compartidas con todas las otras clases tecnológicas.	Redes tecnológicas
		3. Frecuencia de la coocurrencia de patentes	La desviación del número de patentes compartidas por un par de clases tecnológicas del valor esperado bajo la hipótesis de que la asignación de códigos es aleatoria.	Redes tecnológicas

Fuente: Luo y Yan (2016). Traducción propia

La explicación de cada uno de los doce posibles indicadores en detalle excedería el límite en relación con el objetivo de este capítulo. El trabajo de Luo y Yan presenta cada indicador de forma detallada con algunos ejemplos. Además, existe una extensa bibliografía sobre cómo realizar estos cálculos y sus diversas aplicaciones en estudios concretos. Por ejemplo, las medidas de la categoría (1) se utilizan más para medir semejanza *expost*[6] o "grado de relación" (*relatedness*) entre bases de conocimiento o entre clases diferentes. Las medidas de las categorías (2) y (3) tienen una aplicabilidad mayor para el análisis de la diversificación tecnológica de agentes, sectores y países. Las medidas del grupo (4) se utilizan más para analizar las propiedades de las bases de conocimiento y su evolución en el tiempo.

Para mostrar un ejemplo sobre cómo construir la matriz de adyacencia, supongamos que queremos construir el grafo de la base de conocimiento global. Para identificar los nodos, decidimos utilizar el nivel de desagregación de la IPC a tres dígitos (nivel "Clase" $C_1 - C_n; n = 131$), lo que nos llevaría a una matriz de adyacencia de 131 *x* 131. Para estimar los vértices, escogemos la medida 4,3; esto es, queremos contabilizar el número de enlaces entre nodos mediante la frecuencia de las coocurrencias de dominios en la misma patente y ponderar esa frecuencia por la desviación entre el número de patentes compartidas por un par de clases y su valor esperado bajo la hipótesis de que la asignación de códigos es aleatoria. Las bases de datos de patentes nos informan del número de patentes en los que coocurren dos campos técnicos diferentes $\{i \neq j\}$ o dos campos técnicos que son iguales para el mismo nivel de agregación $\{i = j\}$, pero diferentes para niveles de agregación mayores dentro de su clase. Con esta información montamos la matriz de adyacencia no ponderada de un grafo no dirigido (ver Cuadro 5):

6 Esta es una forma de definir semejanza *ex post*, es decir, dos dominios de conocimiento son semejantes si ellos están relacionados mediante citas de patentes.

Cuadro 5. Matriz de adyacencia no ponderada de un grafo no dirigido

	C_1	C_2	...	C_n	$\sum_{j=1}^{n} P_{ij}$
C_1	P_{11}	P_{12}	...	P_{1n}	$\sum_{j=1}^{n} P_{1n} = \sum_{i=1}^{n} P_{n1}$
C_2	P_{21}	P_{22}	...	P_{2n}	$\sum_{j-1}^{n} P_{2n} = \sum_{i-1}^{n} P_{n2}$
...	...	...	...	...	...
C_n	P_{n1}	P_{n2}	...	P_{nn}	$\sum_{j-1}^{n} P_{nj} = \sum_{i-1}^{n} P_{in}$

Fuente: elaboración propia

Ejemplo 1: suponga que tenemos una base de datos con cuatro patentes (A, B, C, D) en la que cada una reporta los siguientes dominios o campos técnicos para un nivel de desagregación de tres dígitos (ver Cuadro 6).

Cuadro 6. Ejemplo de cuatro patentes

Patente	Códigos IPC atribuidos (3 dígitos)		
A	C07	-	-
B	A61	C07	-
C	A61	C12	G01
D	D03(G)	D03(H)	-

Fuente: elaboración propia

La patente A solo reporta un campo técnico, o sea, es una observación del nodo C07 que no la conecta a la red. La patente D reporta dos campos técnicos que son diferentes para un nivel de desagregación de 4 dígitos, pero que son iguales en el nivel de desagregación de 3 dígitos. Consideramos entonces que hay una interacción del campo D03 con él mismo. A partir de esas informaciones, montamos la matriz de adyacencia no ponderada como sigue (ver Cuadro 7):

Cuadro 7. Matriz de adyacencia no ponderada (P_{ij})

	A61	C07	C12	G01	D03	$\sum_j P_{ij}$	P_i
A61	-	1	1	1	-	3	2
C07	1	-	-	-	-	1	2
C12	1	-	-	1	-	2	1
G01	1	-	1	-	-	2	1
D03	-	-	-	-	1	1	1
$\sum_j P_{ij}$	3	1	2	2	1	-	-
P_j	2	2	1	1	1	1	-

Fuente: elaboración propia

Observemos que los valores de la diagonal son aristas cíclicas (lazos) en cuanto que el resto de las celdas son aristas rama. El sumatorio de cada fila sería exactamente igual al sumatorio de la columna correspondiente, dado que la matriz es simétrica.

Una vez determinada la matriz con el número de aristas, definimos la matriz de adyacencia ponderada siguiendo la metodología indicada por Luo y Yan (2016) en la que, suponiendo que la matriz sigue una distribución hipergeométrica, cada elemento a_{ij} sería calculado como:

$$a_{ij} = \frac{P_{ij} - \mu_{ij}}{\sigma_{ij}}$$

Obteniendo la matriz de adyacencia ponderada (ver Cuadro 8):

Cuadro 8. Matriz de adyacencia ponderada

	C_1	C_2	...	C_n
C_1	a_{11}	a_{12}	...	a_{1n}
C_2	a_{21}	a_{22}	...	a_{2n}
...	...	...	...	...
C_n	a_{n1}	a_{n2}	...	a_{nn}

Fuente: elaboración propia

En la que μ_{ij} y σ_{ij} son, respectivamente, la media y la varianza del número esperado de patentes activas en ambas clases $i - j$ calculadas como:

$$\mu_{ij} = \frac{P_i P_j}{T}$$

$$\sigma_{ij} = \mu_{ij}\left(\frac{T-P_i}{T}\right)\left(\frac{T-P_j}{T-1}\right)$$

En la que T representa el número total de patentes que tienen dos o más clases tecnológicas; y P_i, P_j representa, el número total de patentes en que aparecen las clases i y j respectivamente.

Considerando los valores P_i, P_j reportados anteriormente junto con la matriz de adyacencia y que en nuestra base de datos es T = 4, tenemos que (ver Cuadro 9 y Cuadro 10):

Cuadro 9. Desarrollo de la matriz de adyacencia ponderada

$P_i P_j$

	A61	C07	C12	G01	D03
A61	4	4	2	2	2
C07	4	4	2	2	2
C12	2	2	1	1	1
G01	2	2	1	1	1
D03	2	2	1	1	1

M_{ij}

	A61	C07	C12	G01	D03
A61	1	1	0,5	0,5	0,5
C07	1	1	0,5	0,5	0,5
C12	0,5	0,5	0,25	0,25	0,25
G01	0,5	0,5	0,25	0,25	0,25
D03	0,5	0,5	0,25	0,25	0,25

$\left(\frac{T-P_i}{T}\right)\left(\frac{T-P_j}{T-1}\right)$

	A61	C07	C12	G01	D03
A61	0,33	0,33	0,5	0,5	0,5
C07	0,33	0,33	0,5	0,5	0,5
C12	0,5	0,5	0,75	0,75	0,75
G01	0,5	0,5	0,75	0,75	0,75
D03	0,5	0,5	0,75	0,75	0,75

σ_{ij}

	A61	C07	C12	G01	D03
A61	0,33	0,33	0,25	0,25	0,25
C07	0,33	0,33	0,25	0,25	0,25
C12	0,25	0,25	0,19	0,19	0,19
G01	0,25	0,25	0,19	0,19	0,19
D03	0,25	0,25	0,19	0,19	0,19

Fuente: elaboración propia

Matriz de adyacencia ponderada (a_{ij}):

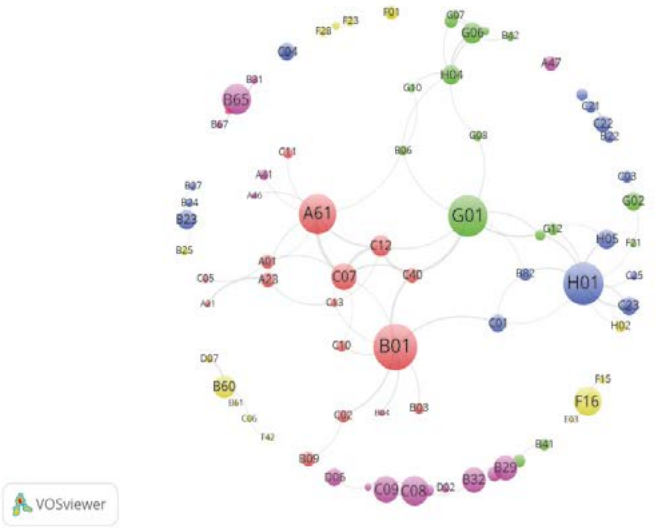
	A61	C07	C12	G01	D03
A61	-3	0	2	2	-2
C07	0	-3	-2	-2	-2
C12	2	-2	-1,33	4	-1,33
G01	2	-2	4	-1,33	-1,33
D03	-2	-2	-1,33	-1,33	4

Fuente: elaboración propia

Propiedades de una base de conocimiento: indicadores

Para construir una base de conocimiento a partir de los datos de patentes utilizaremos los códigos IPC como nodos y la coocurrencia entre ellos en la misma patente para identificar las aristas. En la representación de la red (figura 1), la frecuencia de coocurrencia entre pares del código IPC determina el ancho de las conexiones, mientras que la centralidad del grado refleja la importancia relativa de cada nodo en la red. Los colores clasifican los nodos en grupos.

Figura 1. Representación de una red de conocimiento



Fuente: elaboración propia con algoritmo *Fruchterman-Reingold* en software Pajek (Batagelj y Mrvar, 2016) y visualización con software VOSviewer (Noyons, van Eck y Waltman, 2010)

Las bases de conocimiento tienen cuatro propiedades: variedad, coherencia, distancia cognitiva y convergencia. Estas propiedades describen las características estructurales de las redes y permiten hacer comparaciones entre ellas, por ejemplo, entre las bases formadas por el conjunto de patentes de diferentes regiones o en distintos momentos en el tiempo.

La variedad indica el alcance de la diversificación de una base de conocimiento (Quatraro, 2010; Krafft *et al*., 2011). En la red generada a partir del recuento de coocurrencias entre la IPC, la variedad mide la multiplicidad de pares distintos que la forman. Cuando la base de conocimiento se caracteriza por un número relativamente bajo de aristas, la variedad es baja (Krafft *et al*., 2011). La frecuencia de la coocurrencia es importante para medir variedad, ya que una estructura en la que todos los pares se forman con probabilidades homogéneas genera una situación de mayor variedad que cuando la frecuencia de coocurrencia se concentra en pocos pares. El cálculo del indicador de variedad viene dado por (Frenken y Nuvolari, 2004; Ponds, van Oort y Frenken, 2007):

$$VT \equiv H(X,\ Y) = \sum_i \sum_j p_{ij} \left(\frac{1}{p_{ij}} \right)$$

En el que $p_{ij} = c_{ij}/P$ es la probabilidad de coocurrencia del par de la IPC $i - j$, calculado como la relación entre el número de coocurrencias $i - j$ y las patentes totales del período. La variedad total será mayor cuando menor sea la probabilidad de coocurrencia de un par específico.

Una ventaja en utilizar este indicador es que tiene un carácter multidimensional, es decir, se puede dividir en variedad relacionada y no relacionada para explicar las distintas propiedades de la base de conocimiento en términos de nivel de agregación. La variedad no relacionada (VN) mide la concentración de las coocurrencias $i \neq j$. La variedad relacionada (VR) mide la concentración de coocurrencias $i \neq j$ considerando la pertenencia de los nodos a diferentes dominios (g, z) para niveles más elevados de agregación. Si $i - S_g$ y $j - S_z$, entonces la probabilidad de coocurrencia $i \neq j$ es:

$$p_{gz} = \sum_{i \in S_g} \sum_{j \in S_z} p_{ij}$$

La probabilidad de coocurrencia de los dominios tecnológicos (g, z), con todas las posibles combinaciones de $i \neq j$, se calcula como:

$$H_{gz} = \sum_{i \in S_g} \sum_{j \in S_z} \frac{p_{ij}}{p_{gz}} \left(\frac{1}{\frac{p_{ij}}{p_{gz}}} \right)$$

Por tanto, la variedad relacionada (VR), la no relacionada (VN) y la variedad total (VT) se calculan de la siguiente forma:

$$VR = \sum_{g=1}^{G} \sum_{z=1}^{Z} p_{gz} H_{gz} \quad (1)$$

$$VN \equiv H_Q = \sum_{g=1}^{G} \sum_{z=1}^{Z} p_{gz} log_2 \frac{1}{p_{gz}} \quad (2)$$

$$VT = H_Q + \sum_{g=1}^{G} \sum_{z=1}^{Z} p_{gz} H_{gz} \quad (1 + 2)$$

Un aumento de la variedad no relacionada está asociado con la aparición de discontinuidades en la base de conocimiento, mientras que un aumento de la variedad relacionada indica continuidad en las trayectorias tecnológicas. La relación *VR/VN* indica la relación entre variedad por continuidad y por discontinuidad.

La coherencia se refiere a la interconexión entre dominios tecnológicos (Krafft *et al*., 2011). La coherencia complementa la variedad porque evalúa si la red se diversifica de forma integrada o no. Para medir la coherencia de una base de conocimiento hay tres indicadores: la densidad, la centralidad de grado medio y el coeficiente de agrupamiento medio. Estos indicadores –ya presentados en el segundo apartado–capturan la conectividad de la red. La densidad revela la coherencia de la red porque identifica los aislamientos; la centralidad del grado medio porque mide la capacidad de conexión media entre nodos; y el coeficiente de aglomeración porque captura si el grado de conexión directa entre nodos está más o menos concentrado. Una red con grupos muy concentrados en pocos nodos o con nodos aislados tiende a tener menos coherencia que otra en la que las relaciones son variadas y entre grupos diferentes. La cohesión revela complementariedad del conocimiento, pero también situaciones de *lock-in* cuando ciertos dominios tecnológicos no avanzan

porque dependen del progreso de otros. Normalmente, las bases de conocimiento en las que diferentes grupos están interconectados revelan una mayor probabilidad de formación de nuevas conexiones.

La distancia cognitiva representa el grado de similitud entre nodos de acuerdo con la frecuencia en que se combinan, esto es, si dos nodos se combinan poco es porque, desde un punto de vista cognitivo, son disimilares (Nooteboom, 2000; Breschi, Lissoni y Malerba 2003, Krafft *et al*., 2011). En las redes, el concepto de distancia cognitiva puede analizarse con el uso de medidas ponderadas de los enlaces que dan mayor fuerza a las relaciones más frecuentes, como se vio anteriormente. Una vez que se calcula el índice para todos los pares posibles, se necesita agregarlos para obtener una medida sintética de distancia cognitiva de toda la red. Esa medida puede consistir en indicadores agregados que describen la distancia media entre todos los pares de nodos como el diámetro y la longitud media del camino. En un análisis evolutivo de la base de conocimiento, las medidas de distancia cognitiva son útiles para identificar las discontinuidades en la base, es decir, la aparición de nuevas piezas mal conectadas. La reducción en la distancia cognitiva a través del tiempo indica la acumulación del conocimiento y la madurez de la red (Krafft *et al*., 2011).

En último lugar, la convergencia es la propiedad de la base de conocimiento que atiende la idea de interdisciplinariedad, esto es, la aproximación entre áreas de conocimiento que son disimilares, pero que se complementan para formar nuevas tecnologías. La idea de convergencia está en alguna medida asociada a la propiedad de ubicuidad de algunos dominios tecnológicos, esto es, de formar parte de una amplia variedad de coocurrencias y de integración con dominios tecnológicos disimilares. Una forma de verificar la convergencia en las redes de conocimiento es a partir del análisis de la distribución de nodos entre los grupos formados por algún algoritmo de segmentación. Para formar estos grupos, los algoritmos asignan cada IPC al conjunto en el que el grado de similitud es relativamente mayor dada la frecuencia de coocurrencia. El resultado de la segmentación de la red en grupos permite identificar un proceso de convergencia local que deberá ocurrir entre las IPC que fueron designadas para formar parte del mismo grupo. La convergencia de la red de conocimiento en su conjunto debe analizarse de acuerdo con el número de grupos formados, con el tamaño de los grupos y con la existencia de relaciones entre las IPC que pertenecen a diferentes grupos. Los grupos que están más aislados en la red son aquellos cuyos nodos tienen pocas relaciones fuera de ellos mismos, desarrollan trayectorias específicas y

contribuyen relativamente menos a la aproximación de áreas de conocimiento distintas. Sin embargo, habrá convergencia cuando existen elevadas conexiones entre las IPC que pertenecen a grupos disimilares entre sí.

Una vez identificados los grupos, las IPC se clasifican bajo algún criterio, como, por ejemplo, el paradigma tecnológico, su carácter ubicuo, etc. La agregación de la IPC bajo algún criterio facilita el análisis de convergencia porque permite identificar, por ejemplo, cómo las tecnologías del mismo paradigma se distribuyeron entre grupos en diferentes momentos. La asociación de las IPC con los paradigmas permitiría evaluar si las tecnologías dentro del mismo paradigma tienden a agruparse en un solo grupo compacto o si tienden a relacionarse con diferentes grupos, esto es, convergen con otras tecnologías. Si un conjunto de la IPC permanece en el mismo grupo en diferentes momentos, esto significa que la fuerza de las relaciones entre ellos permanece relativamente estable. Los indicadores de variedad relacionada y no relacionada también pueden ser aplicados en el análisis de la convergencia (Avanci y Urraca-Ruiz, 2021).

Conclusiones y recomendaciones

La aplicación de la teoría de grafos y del análisis de redes al estudio de la complejidad intrínseca a los sistemas de innovación y tecnológicos ha sido ampliamente desarrollada en los años recientes. Este capítulo ha sido apenas una pequeña muestra de esta metodología y cómo pude ser utilizada, concretamente, en el análisis de sistemas tecnológicos. La literatura muestra que los objetivos de estudio en este espacio analítico son múltiples; desde la prospección tecnológica hasta el análisis de trayectorias tecnológicas (por ejemplo, en biotecnología). Además, el método presenta una forma nueva de analizar temas tradicionales en la economía de la innovación y en economía industrial, como la permeabilidad de las tecnologías entre industrias, el análisis de competencias de las empresas e industrias, la diversificación tecnológica, la especialización tecnológica, etc.

Además de las aplicaciones a sistemas tecnológicos, esta metodología permite la representación de redes de sistemas de innovación, por ejemplo, para estudiar fenómenos como la cooperación tecnológica entre empresas, las relaciones universidad-empresa, los flujos internacionales de tecnología entre agentes nacionales y extranjeros o la composición del sistema nacional de innovación a través de multigrafos por agentes

que se interconectan por relaciones de cooperación, de financiación, de mercado, etc.

En América Latina, las aplicaciones de esta metodología utilizando bases de datos de patentes no debería ser un hándicap. Algunos países como Brasil, la Argentina o México cuentan con suficientes observaciones para construir redes y observar su evolución. En los países con menor número de observaciones, es posible estudiar redes a partir de inventores o de citaciones, para estudiar la relación de la base de conocimiento local con la mundial. Las restricciones del número de observaciones solo pueden representar una restricción seria en función de cuál sea el fenómeno analizado. Para el estudio de trayectorias y prospección tecnológica, por ejemplo, se deben utilizar bases agregadas de todo el mundo relativas a las tecnologías o paradigmas específicos de estudio. Finalmente, el estudiante no debe olvidar que el análisis de redes es un método aplicable a la representación de un sistema que, si no es específicamente tecnológico, no necesita del uso de patentes. Las informaciones recogidas a través de datos primarios, relativas a cuestiones territoriales o sectoriales, en las que existe interacción entre agentes y que funcionan como un sistema podrán ser representadas y analizadas también utilizando esta metodología.

Bibliografía

Aho, A.V.; Hopcroft, J. E. y Ullman, J. D. (1998). *Estructuras de datos y algoritmos*. México, DF: Pearson/Addison Wesley.

Avanci, V. y Urraca-Ruiz, A. (2021). "Technology clucles and the evolution of knowledge base complexity since the 1980s". *Revista Brasileira de Inovação*, vol. 20, pp. 1-24. DOI: https://doi.org/10.20396/rbi.v20i00.8655490.

Barabási, A. L. (2016). *Network Science*. Cambridge: Cambridge University Press.

Batagelj, V. y Mrvar, A. (2016). "Analysis and visualization of large networks with program package Pajek". *Complex Adaptive Systems Modeling*, vol. 4, nº 1, pp. 1-8. DOI: 10.1186/s40294-016-0017-8.

Biggs, N. L. (1998). *Matemática discreta*. Barcelona: Vicens Vives.

Bogart, K. P. (2000). *Matemáticas discretas*. México: Limusa.

Bonacich, P. (1972). "Factoring and weighting approaches to status scores and clique identification". *Journal of Mathematical Sociology*, vol. 2, nº 1, pp. 113-120. DOI: https://doi.org/10.1080/0022250X.1972.9989806.

Borgatti, S. P. (2005). "Centrality and network flow". *Social Networks*, vol. 27, nº 1, pp. 55-71. DOI: https://doi.org/10.1016/j.socnet.2004.11.008.

Breschi, S.; Lissoni, F. y Malerba, F. (2003). "Knowledge-relatedness in firm technological diversification". *Research Policy*, vol. 32, pp. 69-87. DOI: https://doi.org/10.1016/S0048-7333(02)00004-5.

Caicedo, A.; Mendez, R. M. y Wagner de García, G. (2010). *Introducción a la teoría de grafos*. Colombia: Elizcom.

Carrington, P. J. y Scott, J. (2011). *The SAGE handbook of social network analysis.* Thousand Oaks, CA: SAGE.

de Heide, M.; Debergh, P.; Som, O.; van de Velde, E. y Wydra, S. (2015). *Key Enabling Technologies (KETs) Observatory*. Unión Europea: European Commission. Disponible en: https://www.eusemiconductors.eu/sites/default/files/uploads/20151216_KETs_Observatory_Second-Report.pdf.

Freeman, L. (1978). "Centrality in social networks conceptual clarification". *Social networks*, vol. 1, nº 3, pp. 215-239.

Frenken, K. y Nuvolari, A. (2004). "Entropy Statistics as a Framework to Analyze Technological Evolution". En Foster, J. y Hölzl, W. (eds.), *Applied evolutionary economics and complex systems*, Cheltenham, Reino Unido/Northampton, Mass.: Edward Elgar.

Grassmann, W. K. y Tremblay, J. P. (2003). *Matemática discreta y lógica*. Madrid: Prentice Hall.

Grimaldi, R. P. (1997). *Matemáticas discreta y combinatoria*. México, DF: Addison Wesley/Iberoamericana.

Jackson, M. O. (2010). *Social and economic networks*. Princeton: Princeton University Press.

Jaffe, A. y Trajtenberg, M. (2002). *Patents, Citations and Innovations: A Window on the Knowledge Economy*. Cambridge, MA: MIT Press.

Kajikawa, Y.; Nakamura, H.; Sakata, I. y Suzuki, S. (2015). "Knowledge combination modeling: The measurement of knowledge similarity between different technological domains". *Technological*

Forecasting and Social Change, vol. 94, pp. 187-201. DOI: https://doi.org/10.1016/j.techfore.2014.09.009.

Kay, L.; Newman, N.; Porter, A. L.; Rafols, I. y Youtie, J. (2014). "Patent overlay mapping: Visualizing technological distance". *Journal of the Association for Information Science and Technology*, vol. 65, n° 12, pp. 2432-2443.

Kogler, D. F.; Leydesdorff, L. y Yan, B. (2017). "Mapping Patent Classifications: Portfolio and Statistical Analysis, and the Comparison of Strengths and Weaknesses". *Scientometrics*, vol. 112, pp. 1573-1591.

Krafft, J.; Quatraro, F. y Saviotti, P. P. (2011). "The knowledge-base evolution in biotechnology: a social network analysis". *Economics of Innovation and New Technology*, vol. 20, n° 5, pp. 445-475.

Kushnir, D.; Leydesdorff, L. y Rafols, I. (2014). "Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC)". *Scientometrics*, vol. 98, pp. 1583-1599.

Luo, J. y Yan, B. (2016). "Measuring technological distance for patent mapping". *Journal of the Association for Information Science and Technology*. DOI: 10.1002/asi.23664.

Maurseth, P. B. y Verspagen, B. (2002). "Knowledge spillovers in Europe: A patent citations analysis". *Scandinavian Journal of Economics*, vol. 104, n° 4, pp. 531-545.

Milesi, D.; Petelski, N. y Verre, V. (2017). "Innovación y uso de patentes en la industria argentina". En Ministerio de Ciencia, Tecnología e Innovación Productiva de Argentina (MINCYT); Ministerio de Trabajo, Empleo y Seguridad Social de Argentina (MTEYSS) y NU. CEPAL, *La encuesta nacional de dinámica de empleo e innovación (ENDEI) como herramienta de análisis: la innovación y el empleo en la industria manufacturera argentina*. Buenos Aires: CEPAL

Nesta, L. y Saviotti, P. P. (2005). "Coherence of the knowledge base and the firm's innovative performance: evidence from the US pharmaceutical industry". *The Journal of Industrial Economics*, vol. 53, n° 1, pp. 123-142.

Nooteboom, B. (2000). "Learning by interaction: absorptive capacity, cognitive distance and governance". *Journal of management and governance*, vol. 4, nº 1, pp. 69-92. DOI: 10.1023/A:1009941416749.

Noyons, E. C. M.; van Eck, N. J. y Waltman, L. (2010). "A unified approach to mapping and clustering of bibliometric networks". *Journal of Informetrics*, vol. 4, nº 4, pp. 629-635. DOI: 10.1016/j.joi.2010.07.002.

Observatoire des Sciences et Techniques (OST) (2010). *Indicateurs de sciences et de technologies*. París: Economica/OST.

Organisation for Economic Co-operation and Development (OECD) (2009). *OECD-Patent statistics manual*. París: OECD. DOI: 10.1787/9789264056442-en.

Ponds, R., Van Oort, F., & Frenken, K. (2007). The geographical and institutional proximity of research collaboration. *Papers in regional science*, 86(3), 423-443.

Quatraro, F. (2010). "Knowledge coherence, variety and economic growth: manufacturing evidence from Italian regions". *Research Policy*, vol. 39, nº 10, pp. 1289-1302.

Rosen, K. H. (1999). *Discrete Mathematics and its Applications*. New York: MacGraw Hill.

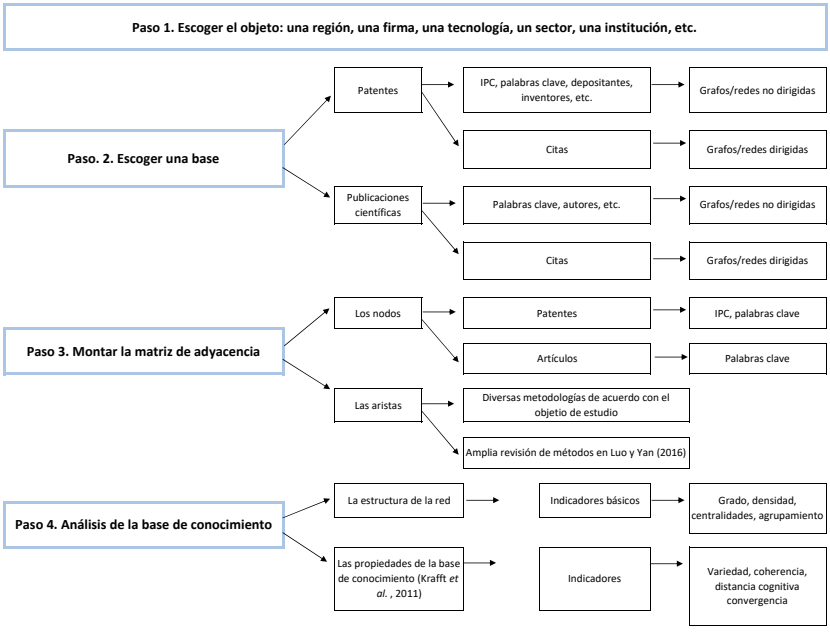
Saviotti, P. P. (2009). "Knowledge networks: structure and dynamics". *Innovation Networks* (pp. 19-41). Berlín, Heidelberg: Springer.

Strogatz, S. H. y Watts, D. J. (1998). "Collective dynamics of 'small-world' networks". *Nature*, vol. 393, nº 6684, pp. 440-442.

World Intellectual Property Organization (WIPO) (2019). *Guide to the international patent classification*. Ginebra: WIPO. DOI: 10.34667/tind.40300.

Anexo

Diagrama. Paso a paso de cómo construir y analizar una base de conocimiento



Capítulo 4
Modelado y simulación de problemas de CTI con dinámica de sistemas[1]

Milton M. Herrera, Mauricio Uriona Maldonado

Introducción

A pesar de que la simulación computarizada ya es un método *mainstream* en las ciencias puras y se aplica desde hace décadas (Belhadi *et al.*, 2022), en las ciencias sociales su inserción es más reciente y menos masificada. Existen por lo menos tres razones por las cuales la simulación puede ser útil a los problemas de las ciencias sociales y en particular para los problemas de CTI en América Latina. Primero, la simulación ayuda a evaluar y diseñar alternativas de solución para un problema, de forma rápida y menos costosa, sin la necesidad de implementar las soluciones en el mundo real y en muchos casos, sin la necesidad excesiva de datos. Segundo, la simulación ayuda a mejorar la comprensión del problema, pues al diseñar y simular un modelo de simulación, el modelador aprende cómo se comporta el problema ante diferentes intervenciones (toma de decisiones). Tercero, la simulación ayuda a definir e identificar las causas de un problema; es decir, contribuye a delimitar el problema que será estudiado. Por lo tanto, facilita la eliminación de factores o elementos que de hecho no afectan al problema y que muchas veces son difíciles de identificar en el contexto real.

1 Los autores agradecen a la Universidad Federal de Santa Catarina y la Universidad Militar Nueva Granada por ser parte parcial de la financiación de la investigación titulada "Diseño de un índice de tecnologías verdes para Latinoamérica" (Grant, IMP-ECO-3402).

Existe una esfera de aplicación que se centra en el estudio de los sistemas dinámicos sociotecnoeconómicos. Esta metodología, también conocida como "dinámica de sistemas" (DS), nació de la adaptación de las técnicas de ingeniería de control para abordar problemas de las ciencias sociales. El origen de la DS se remonta a la década de 1950, cuando Jay W. Forrester del Instituto Tecnológico de Massachusetts (MIT) adaptó los principios de servomecanismos a problemas de dinámica industrial.

La DS se ha utilizado principalmente para modelar sistemas complejos compuestos por múltiples relaciones con el objetivo de crear simulaciones que ayuden a descubrir el impacto de decisiones actuales en el comportamiento futuro del sistema. Un ejemplo es el estudio *Limits to growth* (2004) de Meadows. D, Meadows, D. H. y Randers en el que los autores investigaron la coevolución del crecimiento poblacional, económico y de emisiones de gases de efecto invernadero, ayudando a: 1) descubrir los principales mecanismos de crecimiento, equilibrio y erosión (estancamiento) que impulsan el comportamiento dinámico de los sistemas socioeconómicos, a través de la representación de ciclos de retroalimentación, 2) para reproducir; es decir, simular, el comportamiento dinámico del sistema mediante el uso de ecuaciones diferenciales, 3) entender las implicaciones de los retardos de un sistema en el largo plazo y 4) probar y diseñar alternativas de política que conduzcan a un mejor rendimiento del sistema.

La estructura de este capítulo comienza con esta introducción. A continuación, se presenta la importancia de la simulación en las ciencias sociales. En el tercer apartado se presentan los principales conceptos que implica el uso del método de simulación de dinámica de sistemas (así como el procedimiento para su aplicación). En el siguiente apartado se hace una breve discusión sobre los datos necesarios para la aplicación del método de simulación en CTI, particularmente en América Latina. Finalmente, se presentan las recomendaciones para futuras aplicaciones.

La simulación computarizada en las ciencias sociales

La simulación utiliza modelos formales o matemáticos para representar una porción de la realidad de la cual se tiene interés, por este motivo, se acostumbra decir y aceptar *a priori* que los modelos son simplificaciones imperfectas de la realidad (Downs, Lindsay y Lunn, 2003), y que por lo tanto nacen "equivocados" (Sterman, 2002), aun así, los modelos

bien construidos y seriamente validados pueden ser útiles para apoyar la toma de decisiones.

Existen cuatro metodologías o enfoques de simulación computarizada principales: la simulación de eventos discretos (*discrete event simulation* (DES)), el modelado basado en agentes (*agent-based modeling* (ABM)), la simulación de sistemas dinámicos (*dynamical systems*) y la dinámica de sistemas (*system dynamics* (DS)) (Pidd, 1998; Borshchev, 2013). Varios trabajos se han preocupado por entender las similitudes y diferencias entre estos enfoques (Borshchev, 2013). A continuación, se presentan algunas de las principales conclusiones de estos trabajos y otros relacionados con la conceptualización de los enfoques de simulación.

El primer enfoque de la simulación –DES– se aplica a problemas con un nivel de abstracción medio o bajo, con foco operativo; es decir, utilizado para el mejoramiento del desempeño en procesos o actividades específicas a nivel operativo. Esto significa que es necesario predefinir todos los elementos que conforman el modelo (actividades, tiempos de ejecución, etc.), normalmente haciendo uso de conceptos matemáticos como la "teoría de filas" o la gestión de procesos (Dumas *et al.*, 2013; Barra Montevechi *et al.*, 2022). Un ejemplo de este tipo de simulación son los modelos de tipo *digital twin* utilizados para mejorar la operación de agencias bancarias, definiendo el número ideal de operadores de caja, dada una determinada distribución de llegada de clientes a lo largo de un día, con un tiempo promedio de atención por cliente y con una proporción predefinida de cajeros automáticos y de cajeros humanos.

En la simulación DES lo que importa es determinar lo que ocurre con cada uno de los elementos o instancias que atraviesan el modelo, que para el ejemplo anterior serían los clientes, determinando el tiempo real de servicio por cliente, el tamaño de las filas de espera y los costos para mantener la estructura de operadores.

El segundo enfoque es el de la simulación basada en agentes o ABM, que puede ser usada en problemas que van desde un nivel alto hasta un nivel bajo de abstracción. Los modelos ABM consideran el comportamiento individual de cada agente y a partir de las interacciones entre cientos o miles de agentes generan el comportamiento global del sistema (Borshchev, 2013). Este enfoque es especialmente recomendable en problemas que necesitan modelar el comportamiento individual de un gran número de "agentes" (que pueden ser personas, vehículos, productos, etc.) en el micronivel, para rastrear los resultados de sus interacciones en el macronivel. En la ABM cada elemento o agente (y su comportamiento)

es independiente de los demás, por lo que el resultado de las microinteracciones es difícilmente predecible *a priori*.

A pesar de ser un enfoque relativamente reciente, que ganó popularidad a partir de la década de 1980, ha sido muy utilizado en las ciencias sociales y particularmente en los estudios de innovación para desarrollar modelos evolucionarios (Nelson y Winter, 1982), modelos *history-friendly* (Landini, Lee y Malerba, 2017) y modelos de difusión de innovaciones (Hamacher y Kotthoff, 2022). Sin embargo, la curva de aprendizaje para el desarrollo de modelos ABM es muy lenta, pues no existe un estándar único de programación, ya que existen diferentes lenguajes de programación, dificultando la construcción, migración y adaptación de modelos de un programa informático a otro.

El tercer enfoque es la simulación de los sistemas dinámicos (*dynamical systems*), utilizada en problemas de ingeniería mecánica, eléctrica, química y otras (Ogata, 1998), basada en el análisis de un conjunto de ecuaciones diferenciales que representan el sistema y su complejidad. Los sistemas que con mayor frecuencia se estudian son los físicos; por ejemplo, velocidad, vibración, presión, voltaje y otros que caracterizan un sistema de control (Borshchev, 2013). Por lo tanto, se usan en problemas con baja abstracción y de nivel operativo enfocado en la automatización de los sistemas.

Finalmente, el cuarto enfoque es el de la DS, que difiere del enfoque anterior, principalmente porque se utiliza para resolver problemas en sistemas socioeconómicos (Forrester, 1989). En este sentido, el énfasis de la DS es principalmente en el macronivel y estratégico. Ejemplos de este tipo de enfoque son los modelos de simulación desarrollados para ayudar a verificar el impacto de políticas públicas en países o regiones (Ghaffarzadegan, Lyneis y Richardson, 2011).

Además, los modelos de DS han sido también utilizados en problemas operativos y con bajo nivel de abstracción, en los que los modelos DES han sido predominantes durante varios años (Godinho Filho y Uzsoy, 2009, 2014). En este sentido, los resultados del estudio de Robinson y Tako (2010) demostraron que no existen diferencias significativas en el uso de los modelos DS y DES para problemas operativos, lo que significa que ambos pueden servir para los mismos propósitos.

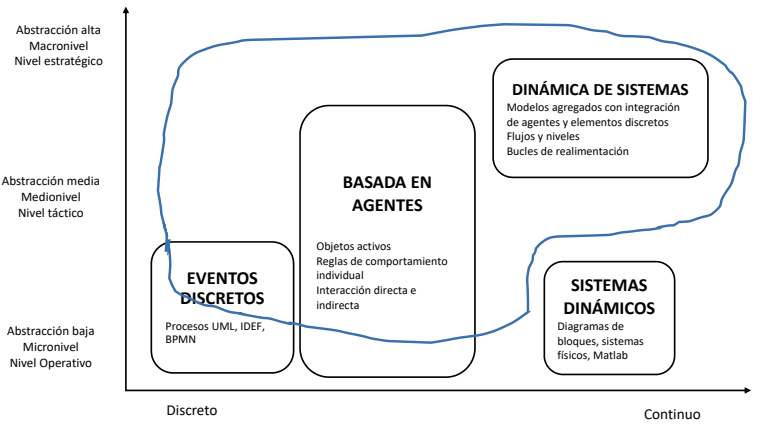
Similar a la ABM, los modelos de DS consideran que los agentes tienen racionalidad limitada y que las interacciones entre ellos (microestructura) generan el comportamiento global del sistema (macrocomportamiento)

desde un punto de vista agregado; es decir, a partir de la agrupación de agentes con características comunes.

De cierta forma, los modelos de simulación basados en DS también pueden ser utilizados en problemas característicos de ABM. Rahmandad y Sterman (2008) compararon los resultados de un modelo ABM y otro modelado con DS para el mismo fenómeno y encontraron que el resultado era muy similar cuando el número de agentes del modelo ABM era significativamente grande, o sea, el modelo ABM tenía una tendencia a llegar al mismo resultado que el modelo basado en DS a medida que el número de agentes aumentaba.

La figura 1 muestra una síntesis de los cuatro enfoques abordados y los localiza de acuerdo con sus características de nivel de abstracción y discretismo/continuismo. Esta figura es útil en la medida que muestra el espectro de aplicación e implementación de los enfoques de los métodos de simulación.

Figura 1. Diferentes enfoques de simulación para las ciencias sociales



Fuente: adaptado de Borshchev (2013)

Descripción del método

Conceptos clave y características principales

Antecedentes de la dinámica de sistemas

En el decenio de 1950, Jay W. Forrester, profesor del MIT, iniciaba una nueva etapa de su carrera dentro de la recién formada Sloan School of Management. El MIT creía que esta escuela de negocios podría apoyarse en la fuerte base técnica de sus profesores y que sería un diferencial de mercado con relación a otras competidoras. En esta línea, Forrester, ingeniero eléctrico que se había dedicado al diseño de radares y microcomputadores, se interesó por el estudio de las empresas y las causas que llevaban al éxito o fracaso de estas.

Basándose en su conocimiento y experiencia previos en la esfera de la ingeniería de control y la retroalimentación, Forrester se dio cuenta de que el concepto de retroalimentación también podría ser aplicado a problemas en las áreas de los negocios.

Esta idea cobró fuerza cuando inició un proyecto con General Electric para explicar la inestabilidad y las fluctuaciones de la cadena de suministro de la planta de Kentucky (Estados Unidos), un problema que no podía explicarse con los métodos tradicionales de gestión de la producción (Forrester, 1989). Después de reunirse varias veces con ejecutivos y personal de planta, Forrester desarrolló el primer modelo de DS que ayudó a identificar que la causa fundamental de las fluctuaciones sufridas por General Electric eran producidas por los retrasos en el flujo de las informaciones generados por las decisiones basadas en informaciones incompletas (Bendoly *et al.*, 2015).

El resultado de esta primera aplicación fue un artículo publicado en 1958. La acogida de utilizar conceptos de retroalimentación en problemas del área de negocios y con simulación computarizada fue tan positiva que años más tarde Forrester publicó el libro titulado *Industrial Dynamics* en el que se explicaba en detalle cada parte del modelo aplicado al caso de General Electric (Forrester, 1958, 1961).

Poco después de la publicación de *Industrial Dynamics*, Forrester comenzó a abordar otros tipos de problemas socioeconómicos más allá del entorno industrial. El estudio que le otorgó mayor reconocimiento mundial fue uno en el que se enfocó en desarrollar un modelo que ayudase a entender las complejas relaciones entre la economía y los recursos naturales y ambientales. Este estudio, denominado "Limits to Growth"

(Forrester, 1973) ganó fama al ser utilizado como uno de los documentos seminales para la elaboración del concepto de *desarrollo sustentable* (Ekins, 1993), el cual generó importantes debates, incluso con investigadores del grupo SPRU de la Universidad de Sussex, donde se discutía el rol de la CTI en los límites del crecimiento (Cole *et al*., 1973). En síntesis, el estudio llamaba la atención a los formuladores de políticas sobre el uso indiscriminado de recursos naturales y los potenciales problemas económicos, sociales y ambientales que podrían ocurrir en las próximas décadas, si los líderes mundiales no diseñaban e implementaban un plan de gestión más eficiente de recursos (Meadows. D, Meadows, D. H. y Randers, 2004).

De esta forma, nacía formalmente la metodología de DS, ganando también interés a escala mundial. En la actualidad, la aplicación de la dinámica de sistemas ha crecido sustancialmente, brindando soluciones a la gestión empresarial, a problemas económicos, ecológicos, fenómenos sociales y de educación. Este crecimiento vertiginoso de las aplicaciones de la DS llevó a la difusión de los conocimientos y aplicaciones de la DS en una revista científica especializada, *System Dynamics Review*. Además, otras actividades como la organización de un congreso internacional anual, para compartir los avances teóricos y prácticos en la materia –International System Dynamics Conference (ISDC)– como también la agrupación de los profesionales e investigadores en una organización internacional –System Dynamics Society– han contribuido al desarrollo de este enfoque de simulación.

Conceptos y premisas del modelado con dinámica de sistemas

Como se ha explicado anteriormente, una primera fuente de los orígenes de la dinámica de sistemas se basa en la ingeniería de control. Una segunda fuente de inspiración es el concepto de sistema complejo. Los sistemas complejos están compuestos por agentes racionales limitados (Sterman, 2000); por patrones acumulativos que producen una causa positiva de refuerzo y equilibrio negativo (Niosi, 2010); debido a la no linealidad y la dinámica de desequilibrio (Niosi, 2004); además, influenciada por los retrasos temporales (Rahmandad y Sterman, 2008).

Así pues, la primera premisa de la DS se refiere a la importancia de la estructura del sistema – los elementos físicos, las reglas de decisión y sus interrelaciones– para explicar el comportamiento de un sistema. Por ejemplo, en el caso expuesto en la obra *Industrial Dynamics* (1961),

Forrester argumenta que la principal causa de la fluctuación de la cadena de suministro de General Electric se debía a la estructura de la toma de decisiones, principalmente relacionada con las decisiones sobre el punto de reposición de las existencias y el desfase entre esas decisiones y los efectos en la cadena.

Posteriormente, varios trabajos se ocuparon de desarrollar nuevas explicaciones sobre el problema de la toma de decisiones en los inventarios y sus fluctuaciones a lo largo de la cadena de suministro. Una obra que merece atención tema este respecto es el artículo de Sterman (1989), en el que el autor explica el fenómeno de las fluctuaciones en la cadena de suministro, argumentando la mala percepción –por parte de los tomadores de decisión– de los fenómenos de acumulación de existencias de productos en proceso, de existencias de productos acabados y de los retrasos en la toma de decisiones. Para ello, Sterman (1989) desarrolló el llamado *beer game* ("juego de la cerveza"), a partir del cual observó un desempeño subóptimo de los jugadores en la posición de gerentes de una cadena de suministro debido a que utilizaban heurísticas simples que no facilitaban el proceso de toma de decisiones, lo que conducía a un resultado subóptimo. En este sentido, Sterman argumenta que la estructura interna del sistema –la cadena de suministro y las reglas de decisión de los gerentes– explica el comportamiento del sistema, el rendimiento caracterizado por períodos de exceso y agotamiento de existencias. Esta visión endógena es particular de la DS.

La segunda premisa de la dinámica de sistemas se relaciona precisamente con el fenómeno de la acumulación, es decir, las respuestas del sistema a las acciones de los responsables de la toma de decisiones se presentan en forma de acumulación –o reducción– de materia, energía o información. En la DS el fenómeno de acumulación se representa mediante niveles, tales como como los gastos en I+D acumulados, el capital intelectual de regiones o países y la capacidad de innovación de países o regiones.

Para la DS, los niveles responden a los cambios en los flujos de entrada y salida. La importancia de los niveles; por lo tanto, es la de ofrecer una visión temporal del sistema en estudio, debido a que los niveles logran captar la "memoria" del sistema; es decir, que, sin cambios en los flujos de entrada y salida, el valor de los niveles permanecerá constante. Un ejemplo de esto es el modelo desarrollado por Fiddaman (2002), en el que el autor demuestra que, incluso con una reducción de las emisiones globales de gases de efecto invernadero (GEI) en la atmósfera, la cantidad

acumulada de CO_2 permanecería sin cambios, porque con este tipo de política, estaría cambiando solo el flujo de entrada (las emisiones de CO_2) y no el flujo de salida (la disipación de CO_2 en la atmósfera), dejando así el stock (el nivel de CO_2 en la atmósfera) constante, demostrando así que es una solución apenas paliativa para el problema del calentamiento global.

De hecho, hay varios trabajos desarrollados en el campo de la DS que demuestran empíricamente la importancia del fenómeno de la acumulación. Este fenómeno está íntimamente ligado con el concepto de la racionalidad limitada, pues los seres humanos no pueden predecir correctamente el comportamiento de sistemas no lineales y complejos, en los que la acumulación existe (Diehl y Sterman, 1995; Weinhardt *et al*., 2015).

La dinámica de sistemas postula que existen, generalmente, más de un nivel dentro de los sistemas socioeconómicos complejos y que estos se influyen mutuamente –a través de sus flujos– de manera dinámica. Esta característica se refiere a la tercera premisa de la dinámica de sistemas, los procesos de *retroalimentación*; es decir, que toda acción o decisión produce una reacción del sistema que recibió esa acción o decisión, alterando el "estado" del sistema en estudio. La figura 2 muestra un ejemplo de la representación utilizada por la DS basada en niveles y flujos para formalizar los conceptos y premisas del modelado.

Figura 2. Ejemplo de representación de niveles y flujos

Fuente: elaboración propia

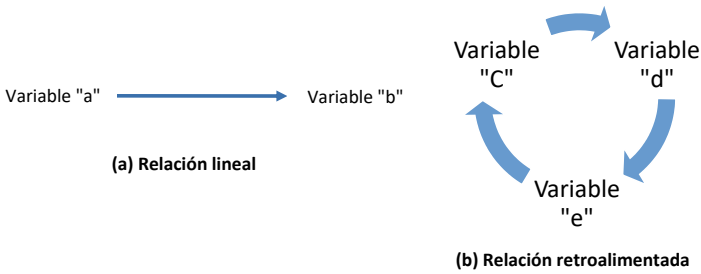
Nota: El nivel está representado como un rectángulo y los flujos como válvulas estilizadas.

Estos procesos de retroalimentación son representados por bucles (o ciclos), que pueden ser negativos o positivos. Un bucle negativo produce una respuesta del sistema en forma de estabilización (o ajuste hacia un valor de equilibrio). Un ejemplo en el área de CTI es el incremento en gastos de I+D que lleva a un incremento en la generación de patentes. De esta manera, un bucle positivo produce una respuesta del sistema que genera un mayor crecimiento. Otro ejemplo también, es la externalidad

de red positiva (Vignieri, 2021), es decir, mientras más usuarios compren un producto, mayor será el efecto multiplicador del boca a boca y mayor la tasa de difusión del producto.

Los sistemas socioeconómicos están formados por más de un bucle de retroalimentación. De hecho, la literatura de DS es clara al afirmar que los comportamientos complejos observados en los sistemas reales son el producto de la interacción de varios bucles de retroalimentación, tanto positivos como negativos. Según Sterman (2006), una característica de los sistemas sociales es que tienen, de la misma manera que los seres vivos, bucles de retroalimentación que autorregulan el comportamiento del sistema en función de respuestas o mecanismos de refuerzo (positivo) y equilibrio (negativo). La falta de conocimiento de los mecanismos de refuerzo y equilibrio más relevantes de un sistema social produce intervenciones que, al ignorarlas, generan las respuestas imprevistas del sistema, conocidas como comportamiento contraintuitivo. La figura 3[2] muestra un esquema de relación lineal y de relación retroalimentada.

Figura 3. Relaciones lineales y retroalimentadas entre variables de un sistema



Fuente: elaboración propia

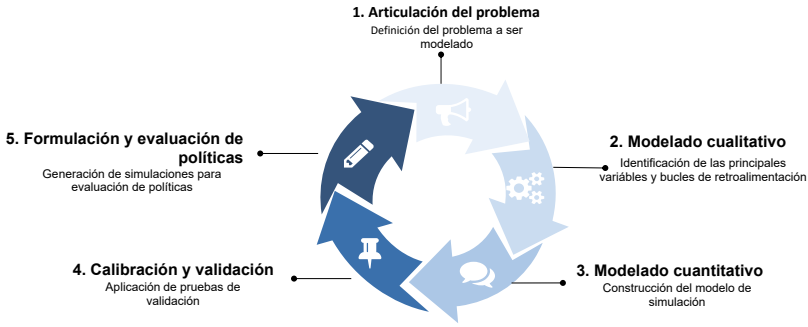
2 Una relación causa-efecto ocurre entre dos variables cuando una afecta el comportamiento de la otra, como se muestra en la figura 3 entre las variables “a” y “b”. Los bucles de retroalimentación se producen cuando hay un bucle cerrado de relaciones (*loop*) entre dos o más variables, como (b) en la figura 3. Este ciclo produce una retroalimentación entre las variables “c”, “d” y “e” que cambia el estado o comportamiento del sistema a lo largo del tiempo.

Finalmente, la cuarta premisa de la DS es el efecto del retardo temporal, o retraso. Este efecto se refiere al hecho de que hay una brecha entre las decisiones y los resultados de estas decisiones, afectando el comportamiento del sistema. Ejemplos de este tipo de fenómeno en la CTI pueden observarse en el retardo existente entre la formulación de una determinada política y los resultados obtenidos por ella (aumento de gastos en la formación de doctores y los resultados regionales o nacionales en términos de innovaciones y tecnologías).

Procedimiento y etapas de aplicación

El modelado en dinámica de sistemas se caracteriza por ser un proceso iterativo y artesanal como se muestra en la figura 4.

Figura 4. Procedimiento de aplicación de la dinámica de sistemas



Fuente: adaptado de Sterman (2000)

Como se observa en la figura 4, el procedimiento está compuesto por cinco etapas que de acuerdo con la propuesta de Sterman (2000) son: 1) articulación (definición) del problema; 2) modelado cualitativo (utilizando diagramas de bucle causal); 3) modelado cuantitativo (utilizando diagramas de niveles y flujo); 4) calibración y validación del modelo y 5) formulación y evaluación de políticas. A continuación, se presentan más detalles de cada etapa, presentando ejemplos en el área de CTI.

Articulación del problema

Esta primera etapa es esencialmente más importante en la DS que en las otras metodologías de simulación. En un estudio realizado por Robinson y Tako (2010), modeladores con más de veinte años de experiencia resaltaron la importancia de esta etapa y que en promedio más de la mitad del tiempo total del proyecto de modelado en DS era dedicado a definir correctamente el problema.

Una de las principales características es que el problema debe ser de tipo dinámico; es decir que debe claramente identificar un comportamiento no deseado (problemático) sobre una línea de tiempo. Producto de esta característica se puede definir la DS como:

> ... un enfoque computarizado para el análisis y diseño de políticas, aplicado a problemas dinámicos presentes en sistemas sociales, de gestión, económicos o ecológicos (literalmente, cualquier sistema dinámico caracterizado por interdependencia, interacción mutua, retroalimentación de informaciones y causalidad circular). (Richardson, 1999)

Esta característica temporal generalmente se representa en la DS como un gráfico conocido como el "modo de referencia" y que sirve para iniciar la identificación de las posibles variables que llevan a ese comportamiento. Este gráfico debe representar la o las variables –las más importantes del problema– que se desea mejorar con la aplicación del modelo de simulación. Ford (2009) presenta algunas reglas para la adecuada articulación del problema que son resumidas en el cuadro 1.

Cuadro 1. Reglas para la articulación del problema

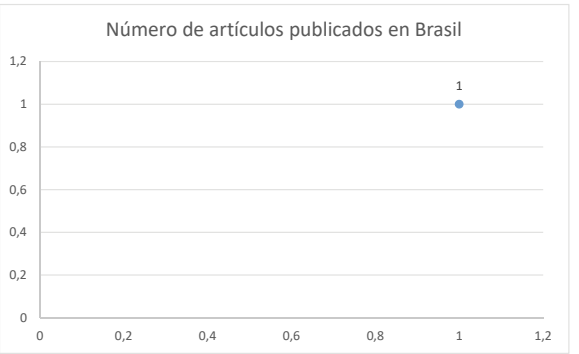
Familiarícese con el problema	Sea específico	Dibuje el problema
El problema debe ser dinámico, o sea, mostrar algún comportamiento variable a lo largo del tiempo considerado	Sea específico para escoger apenas un *problema dinámico*	Utilice dibujos. En el eje "x" el tiempo y en el eje "y" una variable importante. Este gráfico es el *modo de referencia*

Fuente: adaptado de Ford (2009)

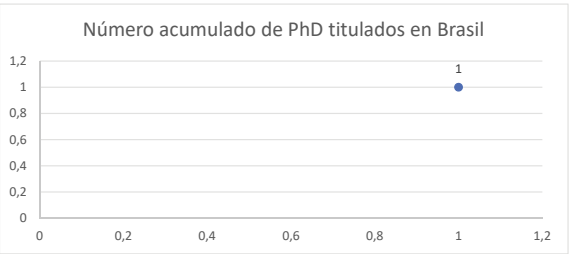
Como se presenta en el cuadro 1, los modos de referencia se trazan normalmente contra el tiempo para facilitar la identificación de la dinámica del comportamiento; es decir, los patrones de cambios del comportamiento en el tiempo de las variables de interés. Los gráficos 1 muestran dos ejemplos de series temporales que pueden servir para entender el concepto de "modos de referencia" en el área de CTI. Estos ejemplos representan un estudio interesado en entender por qué las publicaciones de artículos científicos se estabilizaron cuando el número de doctores titulados continuó creciendo en el mismo período (asumiendo como premisa que por lo menos una parte importante de esas publicaciones es realizada por doctores).

Gráficos 1. Ejemplos de "modos de referencia": número de doctores y de artículos publicados en Brasil (2010-2016)

	Artículos
2010	127860
2011	148435
2012	171969
2013	200029
2014	213731
2015	214633
2016	218588



Año	PhD
2010	11210
2011	12267
2012	13879
2013	15610
2014	17262
2015	18983
2016	20599



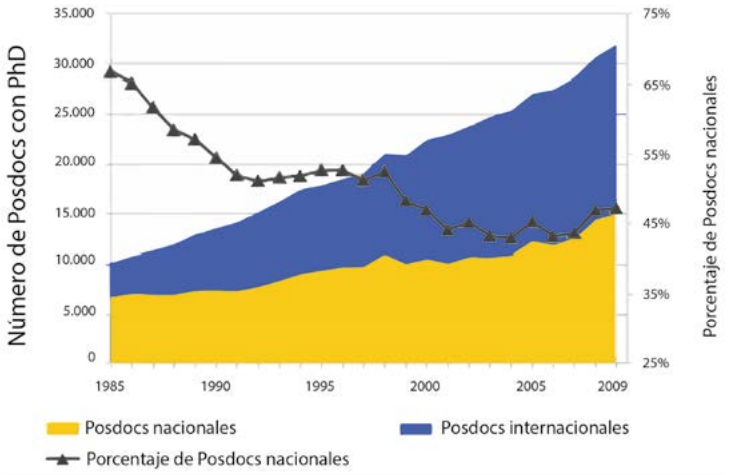
Fuente: CAPES (2017)

Un estudio similar fue realizado por Desai, Ghaffarzadegan y Hawley (2014), en el que los autores intentan identificar cuáles políticas de CTI evitarían una fuga de cerebros de estudiantes de posdoctorado en Estados Unidos y al mismo tiempo una mayor asimilación de este personal calificado en la Universidades de ese país, en el área biomédica. Para esto,

los autores utilizan como modo de referencia datos sobre el número de estudiantes de posdoctorado nacionales y extranjeros para el período 1985-2009, como se ilustra en el gráfico 2. En este gráfico se puede apreciar que el número acumulado de estudiantes de posdoctorado viene aumentando en Estados Unidos, en el período analizado, y que el mayor crecimiento se dio con los estudiantes de origen extranjero.

De esta forma, el problema definido por Desai, Ghaffarzadegan y Hawley es "investigar las tendencias dinámicas de estudiantes de posdoctorado nacionales y extranjeros y desarrollar una herramienta de apoyo a la política que ayude a analizar qué medidas pueden ser tomadas por el gobierno, para afectar estas tendencias en el futuro".

Gráfico 2. Modo de referencia para el número de PhD nacionales y extranjeros



Fuente: Desai, Ghaffarzadegan y Hawley (2014)

Finalmente, es bueno resaltar que los modos de referencia, así como el proceso de diagnóstico y comprensión del sistema real, se alimentan de diversas fuentes de información, documentos, entrevistas, datos estadísticos y otros que incluyen la experiencia de expertos en el sistema socioeconómico en estudio.

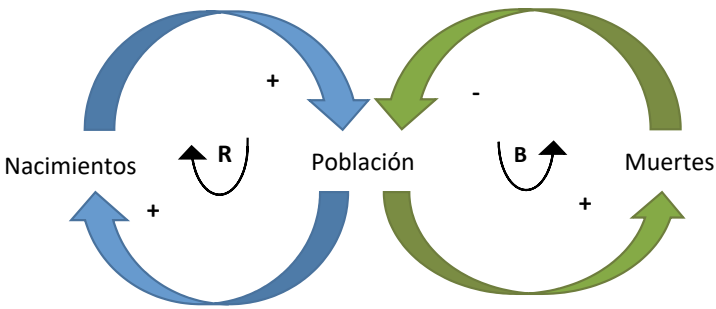
Modelado cualitativo/causal loop diagram

La siguiente etapa es establecer o formular las llamadas "hipótesis del comportamiento dinámico", es decir, la formulación de las posibles relaciones y estructuras de retroalimentación que se consideran las causas del comportamiento observado en los modos de referencia. Para ello, se desarrolla o construye un diagrama de bucle causal (*causal loop diagrams* (CLD)) que incluye los principales bucles de retroalimentación, que se considera, son los responsables por ese comportamiento.

Los CLD tienen como objetivo representar las relaciones causales y no lineales entre las variables del sistema, así como los bucles de retroalimentación y retrasos en el sistema. Un ejemplo de la notación CLD utilizada se muestra en la figura 5 para un sistema de crecimiento de una población. En el ejemplo, las relaciones entre las variables se observan mediante flechas que definen la dirección del efecto, así como la polaridad del efecto, siendo este último positivo (+) o negativo (-).

Por un lado, la polaridad positiva significa que, si hay un aumento en la causa, el efecto también aumentará, por lo que, si la causa disminuye, el efecto también disminuirá. En el caso del ejemplo de la figura 5, un aumento en el número de nacimientos produciría un aumento en la población.

Figura 5. Ejemplo de un modelo de bucles



Fuente: elaboración propia

Nota: Polaridades positivas denotadas por flechas azules y polaridades negativas denotadas por flechas verdes.

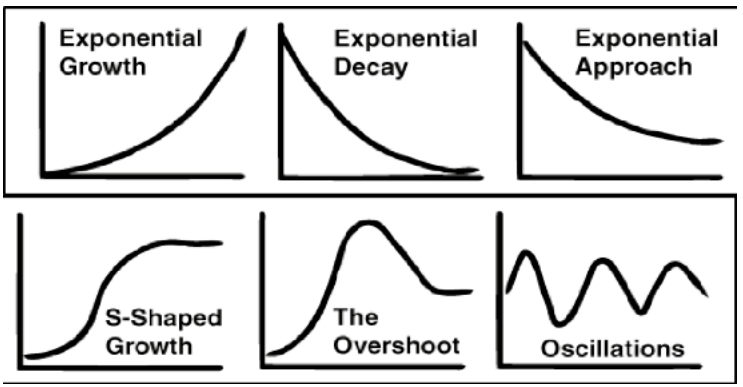
Por otro lado, la polaridad negativa significa que, si hay un aumento en la causa, el efecto disminuirá, o, si hay una disminución en la causa, el

efecto aumentará. En este sentido, un aumento en el número de muertes produciría una disminución de la población.

Según la polaridad de las relaciones, los CLD también representan la polaridad de los bucles de retroalimentación. Por lo tanto, el ciclo de retroalimentación se refuerza cuando la polaridad de las relaciones insertadas en ella es predominantemente positiva (representada por la letra R); y el ciclo será de equilibrio o estabilización cuando la polaridad de sus relaciones sea predominantemente negativa (representada por la letra B).

Dependiendo del problema a ser estudiado, los CLD pueden ser bastante complejos. Retornando al ejemplo de los doctores titulados y los artículos publicados en Brasil, se había comprobado a través de los modos de referencia que mientras el número de doctores aumentaba, el número de publicaciones se estabilizaba. La primera hipótesis dinámica en este caso sería que esa estabilización es causada por una saturación en el crecimiento de las publicaciones, produciendo un comportamiento similar al de una población que alcanza su límite poblacional, también conocido como "crecimiento en forma S". Este procedimiento tiene como objetivo aproximar el "modo de referencia" observado en el caso de estudio a una versión estilizada de las seis versiones más comunes en las aplicaciones de DS. Ford (2009) explica que estas versiones son: crecimiento exponencial, decrecimiento exponencial, aproximación exponencial, crecimiento en forma S, crecimiento excedido y oscilaciones, como muestra la figura 6.

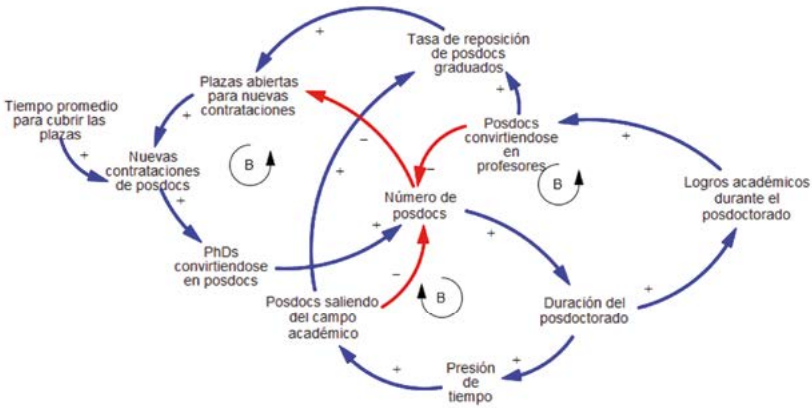
Figura 6. Principales modos de referencias estilizados para proposición de las hipótesis dinámicas



Fuente: Ford (2009)

Ya para el ejemplo de los estudiantes de posdoctorado en Estados Unidos, los autores Desai, Ghaffarzadegan y Hawley (2014) presentan los CLD que se muestran en la figura 7 (adaptada por nosotros para simplificar su comprensión).

Figura 7. CLD simplificados representando los principales bucles que impactan en el crecimiento de posdocs en Estados Unidos



Fuente: adaptado de Desai, Ghaffarzadegan y Hawley (2014)

La figura 7 representa tres bucles de retroalimentación estabilizadora. Dos de ellos representan la relación entre la duración del posdoctorado con las decisiones que llevan a esos estudiantes a salir del sector académico o a obtener cargos docentes en universidades. Mientras que el tercer bucle representa la gestión de contrataciones de nuevos posdocs, a partir del cálculo de plazas abiertas.

Dependiendo de los valores específicos de cada una de las variables, algunos bucles serán más dominantes que otros, dando lugar a comportamientos dinámicos. El problema es que, con un nivel de complejidad similar al del ejemplo anterior, ya es imposible predecir, *a priori*, el comportamiento del sistema, precisamente por el número y la influencia específica de cada uno de los bucles de retroalimentación.

Tampoco permite representar los efectos de una variable sobre otra; por ejemplo, si el "tiempo promedio para cubrir plazas" es muy alto, por

falta de personas interesadas en este tipo de trabajo, o por falta de oferta de doctores, ciertamente habrá un atraso en cubrir las plazas y las contrataciones demorarán más de lo que sería considerado. El problema es que aun con esta hipótesis razonable, es imposible cuantificar el impacto en términos de tiempo o costo.

Por este motivo, se recurre a las herramientas de análisis cuantitativo; en otras palabras, al modelo de simulación en el lenguaje de niveles y de flujos.

Modelado cuantitativo/diagrama de niveles y flujos

El próximo paso en el proceso de modelado es la formulación de un modelo de simulación representado en el lenguaje de los niveles y flujos, que debe seguir la estructura de los CLD construidos. El diagrama de niveles y flujos se emplea para desarrollar la simulación de los efectos de la retroalimentación, los retrasos y los cambios a lo largo del tiempo, a través de la resolución de un sistema de ecuaciones integrales. Cada nivel se puede representar como una ecuación integral de la forma:

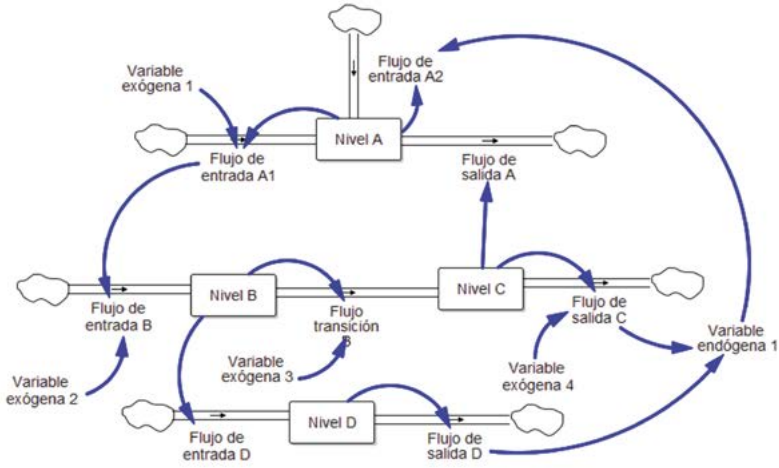
(1)

$$S = \int_a^b f_i(t)dt - f_o(t)dt + S_{t-1}$$

En la que: S_t = nivel "S" en el tiempo *t*; f_i = sumatoria de flujos de entrada; f_o = sumatoria de flujos de salida.

La figura 8 presenta un modelo de simulación con niveles y flujos. Esta figura muestra la relación entre las variables de flujo, los niveles y las variables auxiliares del sistema. Estas últimas suelen ser de dos tipos: valores constantes que permiten modelar las variaciones del sistema que son denominados exógenos y variables endógenas, o sea, calculadas internamente a partir de la interacción con los otros elementos del modelo.

Figura 8. Ejemplo de un modelo de simulación utilizando la notación de niveles y flujos



Fuente: elaboración propia

Matemáticamente, los niveles se relacionan con flujos que siguen la estructura de la ecuación 1. Por lo tanto, el orden del sistema de ecuaciones integrales dependerá del número de niveles en el sistema. Por ejemplo, el sistema de la figura 8 tiene cuatro niveles (A, B, C y D) conectados por flujos, variables exógenas y endógenas. Este sistema puede ser, por lo tanto, representado por un sistema de ecuaciones integrales de cuarto orden.

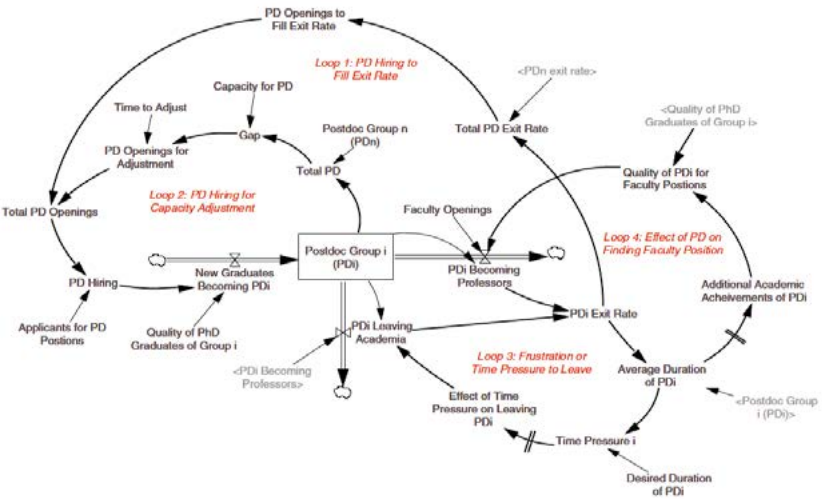
Un aspecto que debe ser resaltado en la figura 8 y que es central en el método de DS es la importancia de representar las relaciones internas entre las variables del sistema. Esto significa que el comportamiento dinámico que se observa en el sistema real debe ser reproducido (simulado) en el modelo de manera endógena.

La importancia de la visión endógena en dinámica de sistemas es tal que ha merecido una posición preponderante a la hora de evaluar, por ejemplo, la calidad de un determinado modelo (se verá más de esto cuando se aborde el tema de la validación).

En la figura 9 se puede apreciar el modelo de niveles y flujos de Desai, Ghaffarzadegan y Hawley (2014), en el que se incluye apenas un único nivel; sin embargo, fuertemente influenciado por más variables endógenas

que por exógenas.[3] En este modelo, el nivel "Postdoc Group i" representa en la realidad dos grupos distintos (o sea, dos niveles independientes), uno de estudiantes de posdoctorado nacionales y otro de extranjeros. De la misma forma, pueden representarse diversos subgrupos de una población, en otras palabras, desagregar un modelo hasta el grado que sea necesario para poder capturar las dinámicas importantes en el sistema, lo cual se aproximaría a un modelo de simulación basado en agentes (ABM).

Figura 9. Modelo de niveles y flujos para el ejemplo del crecimiento de posdocs en Estados Unidos



Fuente: adaptado de Desai, Ghaffarzadegan y Hawley (2014)

Calibración y validación del modelo

En esta etapa se realizan dos actividades diferentes, la primera es la calibración y la segunda es la validación del modelo. La primera actividad se refiere al ajuste de los parámetros del modelo (variables exógenas) a

3 Las variables exógenas son aquellas que no tienen flechas ingresando en ellas, como, por ejemplo, en la figura 9 la variable "Applicants for PD positions", y generalmente son valores fijos o constantes, provenientes de datos históricos.

valores históricos observados en el sistema real. Estos valores históricos pueden provenir de varias fuentes, como sistemas de información, reportes técnicos o científicos, microdatos, estadística descriptiva o inferencial, entre otras. Por ejemplo, para el caso de la figura 9, la variable "Time to Adjust" por ser exógena requerirá de un determinado valor numérico. Este valor puede venir de las informaciones recopiladas en las bases de datos de las agencias de fomento para identificar el tiempo promedio que se requiere para poner a disposición efectivamente una plaza. En otros casos, el valor de las variables exógenas vendrá del ajuste o la calibración realizada por los programas informáticos empleados en el modelado de DS. En estos procedimientos, los programas informáticos buscan valores dentro de un determinado rango, que ayuden a que el modelo se aproxime numéricamente a los datos reales.

En la práctica, la calibración del modelo se realiza junto con la validación, pues a medida que se verifica si los valores calibrados se ajustan adecuadamente al comportamiento del sistema real, se hacen también cambios o mejoras en la estructura del modelo (verificando las relaciones, aumentando o disminuyendo variables, alterando niveles o flujos y verificando las ecuaciones).

La segunda actividad es la de validación o verificación del modelo. Las pruebas de validación en la DS no son apenas estadísticas, porque la cantidad de variables e interrelaciones dificultan el análisis tradicional (Barlas, 1996). Para esto, existen varios tipos de pruebas diferentes que básicamente se pueden clasificar en pruebas estructurales y pruebas de comportamiento (Forrester y Senge, 1980).

Según Seong y Qudrat-Ullah (2010), las pruebas estructurales se refieren a la fiabilidad de la estructura del modelo, ya que confirman o no si la estructura se ha identificado correctamente, en función de la identificación de las variables y parámetros que la componen. En este sentido, las pruebas sugeridas por Forrester y Senge (1980), Sterman (2000), Seong y Qudrat-Ullah (2010) son las siguientes:

- E1, definición de las fronteras del modelo: esta prueba consiste en considerar las relaciones estructurales necesarias para satisfacer el propósito del modelo, verificando si la elección de variables endógenas, exógenas y excluidas tiene sentido. Está prueba puede ayudar a determinar el alcance del modelado (ver el apartado tres).

- E2, verificación de la estructura: esta prueba compara la estructura del modelo con la estructura real del sistema. Por lo tanto, la estructura del modelo no debe contradecir el conocimiento sobre el sistema real. Por ejemplo, en el caso del crecimiento de posdocs en Estados Unidos, la estructura de niveles y flujos de la figura 9 no contradice lo que se conoce sobre la generación de plazas, contrataciones y finalizaciones de trabajos posdoctorales.
- E3, verificación de los parámetros: el propósito de esta prueba es verificar la validez de los parámetros o constantes utilizados en el modelo y compararlos con el conocimiento real de estos, para determinar si corresponden conceptual y numéricamente a la realidad, y para determinar también si se han estimado adecuadamente (muy relacionado con la propia calibración).
- E4, consistencia de unidades y dimensiones: esta prueba verifica si las dimensiones o unidades utilizadas en las ecuaciones son coherentes con las de las variables, parámetros y constantes que las componen.
- E5, condiciones extremas: esta prueba verifica si el modelo se comporta de manera irracional cuando se establecen valores extremos para parámetros o variables.

Aunque las pruebas de estructura sobre el modelo son necesarias, no son suficientes. Por lo tanto, las pruebas del comportamiento del sistema sirven para verificar si el comportamiento generado por el modelo se compara con el comportamiento observado o esperado del sistema real. Las pruebas sugeridas por Forrester y Senge (1980) son las siguientes:

- C1, error de integración: esta prueba verifica un cambio en el comportamiento del sistema cuando el paso de integración (*integration step*) o el método de integración cambia (relativo al método de integración numérico seleccionado por defecto en el software).
- C2, reproducción del comportamiento: el objetivo de esta prueba es verificar si el comportamiento obtenido en el modelo es similar al comportamiento observado en el sistema real. Esta prueba utiliza indicadores estadísticos tradicionales para realizar esa

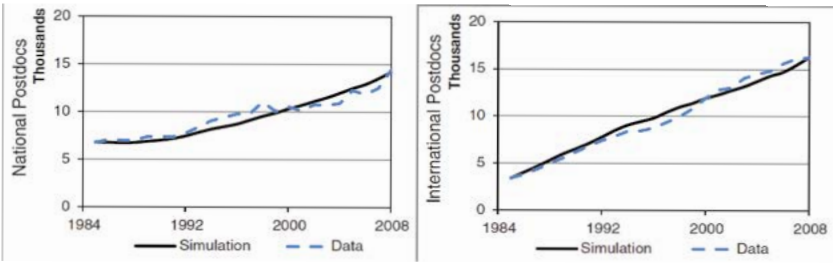
comparación, como por ejemplo R^2, *root mean square error* (RMSE), *mean absolute deviation* (MAD) y otros.

- C3, anomalía de comportamiento: esta prueba se utiliza implícitamente en la construcción del modelo y tiene como objetivo identificar si hay comportamientos anómalos que son producto de suposiciones erróneas en la estructura del modelo.
- C4, reproducibilidad: esta prueba verifica la escalabilidad del modelo para otras realidades similares o sistemas reales.
- C5, comportamiento inesperado: esta prueba verifica si los comportamientos inesperados –producto de las simulaciones– son producto de fallas en la construcción de modelos o si realmente representan comportamientos que son fieles a la realidad y que se han pasado por alto en el sistema real. Si esta es la segunda opción, la prueba demuestra la utilidad práctica del modelo al presentar comportamientos que no se tuvieron en cuenta en el sistema real.

Es importante resaltar que independientemente del gran número de pruebas, su uso depende en gran medida del sistema en estudio, lo que permite dejar de lado algunas pruebas que no contribuyan significativamente a la validez del modelo en cuestión (Seong y Qudrat-Ullah, 2010).

Una forma tradicional de representar la validación numérica es a través de la prueba de reproducción de comportamiento. Los gráficos 3 son un ejemplo tomado de Desai, Ghaffarzadegan y Hawley (2014), en este trabajo se comparan gráficamente los resultados de la simulación con los datos históricos para el número de estudiantes de posdoctorado nacionales y extranjeros.

Gráficos 3. Ejemplo de reproducción de comportamiento



Fuente: Desai, Ghaffarzadegan y Hawley (2014)

Formulación y evaluación de políticas

Es la última y la más importante de las etapas del modelado con dinámica de sistemas. La formulación y evaluación de políticas se refiere a verificar cuál será el comportamiento futuro del sistema si la política actual continúa en ejercicio y así también cuál será el comportamiento si se realizaran cambios en la política actual. En otras palabras, el principal uso de modelos de dinámica de sistemas es el de realizar análisis del tipo qué pasa si... (*what if...*).

Aquí es importante definir lo que se entiende por "política" en la DS. Básicamente, son reglas que determinan acciones o decisiones destinadas para alcanzar un determinado resultado o meta (Forrester, 1992). Se asume, en la DS, que estas políticas son posibles de alteración o, por lo menos, que su alteración es factible. Por ejemplo, en un modelo construido para determinar los niveles óptimos de I+D, la política de ser posible de alteración sería justamente la de inversión en I+D, y los valores de inversión serían los escenarios a ser evaluados por la simulación.

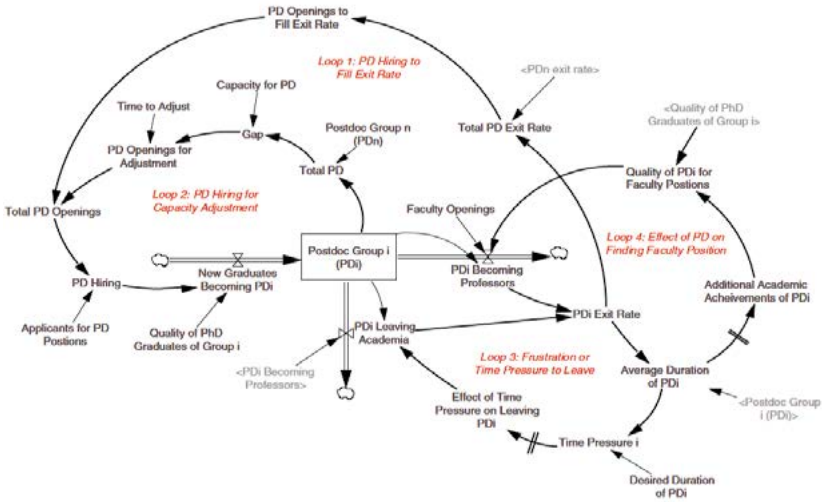
Como se puede observar, el concepto de "política" adoptado por la DS es bastante flexible, por lo que las leyes pueden ser consideradas políticas (por ejemplo, inversión forzosa del 1% de la facturación de las empresas brasileñas del sector de energía en actividades de I+D), las estrategias pueden ser consideradas políticas (por ejemplo, decidir entre una publicidad de masa y una publicidad de nicho) y las decisiones también pueden ser consideradas políticas (por ejemplo, decisión de un consumidor en comprar un vehículo eléctrico versus comprar un vehículo a biocombustible).

Para el caso de Desai, Ghaffarzadegan y Hawley (2014), fueron evaluadas cuatro políticas: 1) limitar el financiamiento de posdocs para cuatro años; 2) limitar el financiamiento de posdocs para dos años; 3) aumentar las contrataciones de profesores en Estados Unidos y 4) incrementar la calidad de la educación y formación de estudiantes de doctorado en Estados Unidos.

Dada la importancia de esta etapa, explicaremos brevemente las características de las políticas anteriores y cómo ellas se relacionan con el problema dinámico definido en el tercer apartado. Primeramente, recordemos el objetivo del modelo definido como: "investigar las tendencias dinámicas de estudiantes de posdoctorado nacionales y extranjeros y desarrollar una herramienta de apoyo a la política que ayude a analizar qué medidas pueden ser tomadas por el gobierno, para afectar estas tendencias en el futuro". Relativo a este objetivo, el problema dinámico estaba enfocado en el desbalance entre posdocs nacionales y extranjeros y en cómo se podría balancear esta relación, aumentando el número de posdocs nacionales.

Sobre las políticas 1 y 2: la idea de limitar el financiamiento de posdocs para cuatro y dos años tendría un impacto en el tiempo total por el que los estudiantes de posdoctorado pueden mantenerse en este cargo, forzando a que el flujo de salida sea mayor; por tanto, a que los posdocs busquen con mayor intensidad una posición en el mercado académico o en el sector productivo (área A en la figura 10). La política 3 busca incrementar las contrataciones de profesores en las universidades americanas. El efecto de esta política sería acelerar el flujo de salida de posdocs, específicamente para que puedan ocupar más cargos permanentes (área B en la figura 10). Finalmente, la política 4 busca incrementar el flujo de entrada (región C en la figura 10) de nuevos posdocs.

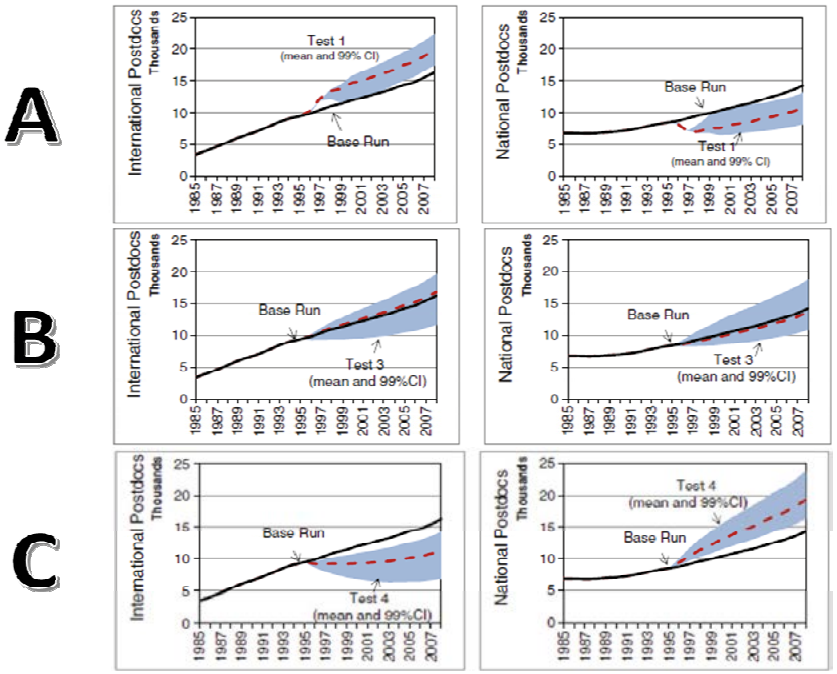
Figura 10. Ejemplo de localización de los impactos de cada política a ser evaluada en el modelo de niveles y flujos



Fuente: Desai, Ghaffarzadegan y Hawley (2014)

A partir de la formalización de las políticas a ser evaluadas, se inician las corridas de simulación, realizando una a una las alteraciones propuestas en las políticas 1, 2, 3 y 4. El resultado también se presenta en formato de gráfico, como muestran los gráficos 4, en los que se han separado los resultados con las mismas letras, A, B y C, de la figura 10. Como se puede observar, el modelo permite evaluar cuál sería el comportamiento de crecimiento de estudiantes de posdoctorado nacionales y extranjeros si cada una de las políticas hubiera sido implementada. En el caso de las políticas 1 y 2, para fines de simplificación se mostrará apenas el resultado de la política 1 en el panel A, en el que se muestra que la limitación de cuatro años para los cursos posdoctorales de hecho tiene un efecto inesperado (ver el ítem C5 en el tercer apartado) con relación al número de estudiantes nacionales que disminuye mientras el número de estudiantes internacionales aumenta.

Gráficos 4. Ejemplo de resultados del análisis de políticas



Fuente: Desai, Ghaffarzadegan y Hawley (2014)

En el panel A, el resultado es inesperado porque lo que se deseaba era incrementar el número de posdocs nacionales. En realidad, conforme explican Desai, Ghaffarzadegan y Hawley (2014), los estudiantes nacionales acaban siendo incentivados a salir más rápido de estos cargos pues históricamente son el grupo que más años pasa en esta posición. Como los proyectos y financiamientos no paran, los investigadores senior son forzados a contratar cada vez más estudiantes extranjeros (por lo que el número de extranjeros aumenta como se observa en el panel A de la izquierda).

La política 3, crear más plazas para profesores en las universidades; de hecho, tiene un efecto positivo para ambos grupos, nacionales y extranjeros, por lo que tampoco se considera como una medida adecuada para alcanzar el objetivo del modelo. La política 4, que tenía impacto en mejorar la calidad de la educación doctoral, sí tiene un efecto positivo en

el grupo deseado, aumentando la proporción de estudiantes nacionales más de lo que lo lograron las otras simulaciones.

Las conclusiones de este estudio, por lo tanto, serían que la política de CTI deben enfocarse en mejorar la educación doctoral y preparar mejor a los recién doctorados, para que puedan ser más competitivos en los procesos de selección de posdoctores. Naturalmente, el modelo no tiene en consideración los diferentes mecanismos que deberían ser implementados para que esta política fuese puesta en práctica (por ejemplo, debate y aprobación de alteraciones en la legislación que permitan mayores inversiones en la educación de los doctores). Sin embargo, el modelo muestra cuáles serían las consecuencias si cada una de esas políticas fuese implementada y ayuda a escoger la más exitosa.

Este ejemplo muestra de una forma secuencial, cómo fueron elaborados todos los pasos en la formulación de un modelo de DS y también la forma como son generalmente usados.

Implicancias para América Latina y el Caribe

La brecha entre el diseño e implementación de políticas que contribuyan al desarrollo de los países en América Latina se ha convertido en un desafío y oportunidad para los tomadores de decisión de política, investigadores y empresarios. El diseño de políticas encaminadas a estimular las actividades de CTI requiere de un análisis sistémico, que permita entender a largo plazo los efectos en los actores del sistema de CTI (Raven y Walrave, 2016). Una mirada desde el modelado y simulación de la dinámica de sistemas contribuye al diseño de alternativas que contempla las dificultades de la implementación de políticas provocadas por las demoras propias del sistema (Cosenz, Dyner y Herrera, 2019).

Estudios previos han utilizado la DS para evaluar políticas relacionadas con la transición de tecnologías limpias en América latina (Dyner y Zuluaga, 2007; Dyner, Franco y Jimenez, 2016). Las alternativas tecnológicas y los desafíos en términos ambientales, sociales y políticos en América Latina requieren de un análisis y modelado que contemple las interacciones sistémicas, demoras e implicaciones a largo plazo de las decisiones políticas. En este sentido, la DS es una metodología útil para el análisis de los problemas propios de América Latina concernientes con el diseño, desarrollo e implementación de las políticas de CTI.

Una mirada de la sostenibilidad de los sistemas de CTI en América Latina es un reto para los decisores de política, investigadores y empresarios. La sostenibilidad de las políticas de CTI contempla un análisis de la gestión tecnológica y de innovación que pueda apoyarse en herramientas de modelado, que permitan entender las interacciones y cambios a largo plazo (Raven y Walrave, 2016). A medida que los países en América Latina mejoren sus prácticas y procesos de toma de decisión, el bienestar de la población mejorará considerablemente. Para ello, el diseño y la implementación de políticas de CTI deben contemplar una mirada para diferentes escenarios que contribuya a entender mejor los cambios de las políticas y las alternativas de mejoramiento.

A partir del punto de vista de la aplicación práctica, la DS es un método de fácil acceso. Por ejemplo, existen softwares gratuitos para académicos, como el Vensim PLE o el AnyLogic PLE, que necesitan apenas de una comprobación del vínculo académico. En términos de la aplicación de la DS, no existen grandes obstáculos para académicos en América Latina, pues, el acceso a datos numéricos no es necesariamente una limitación del método (como se verá en las reflexiones finales).

Desde el punto de vista de las aplicaciones en el área de CTI, existen varias ya publicadas. El estudio de Grobbelaar y Uriona-Maldonado (2018) identificó aplicaciones de dinámica de sistemas en las siguientes áreas: 1) I+D organizativo; 2) difusión tecnológica; 3) estudios nacionales y sectoriales de CTI y 4) estudios en aglomerados productivos e industriales.

Reflexiones finales

En esta última sección se presentarán algunas reflexiones finales sobre los recursos necesarios para la aplicación de la DS, incluyendo: 1) software, 2) datos e informaciones y 3) materiales bibliográficos.

Como se vio en el tercer apartado, la DS utiliza sistemas de ecuaciones integrales para representar sistemas complejos que no pueden ser resueltos analíticamente. Esto significa que existe la necesidad de software específico para realizar esta resolución y por lo tanto conseguir correr las simulaciones.

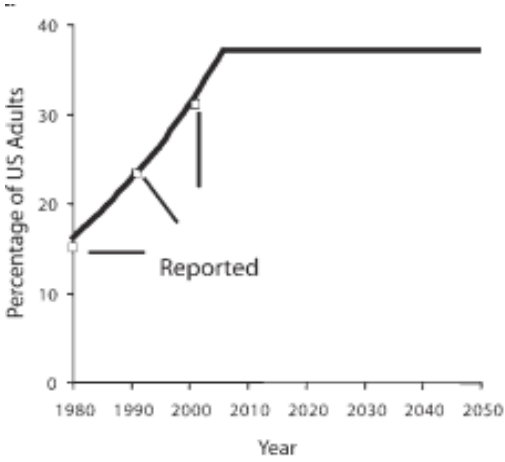
En la actualidad, softwares como Stella Architect de isee systems[4] y Vensim de Ventana Systems[5] facilitan la construcción, interpretación y análisis de los modelos, dejando la resolución numérica de las ecuaciones integrales detrás de interfaces gráficas amigables. Así, un recurso necesario es el acceso a estos softwares y, como se mencionó en la tercera parte, existen softwares gratuitos para académicos, siendo los más conocidos, el Vensim versión PLE y el AnyLogic, versión PLE. Existen incluso paquetes en R (Duggan, 2016) y Python (Houghton *et al.*, 2022) que fueron desarrollados y/o adaptados para que puedan ejecutar los modelos de simulación de DS.

Con relación a los datos e informaciones necesarios, los modelos de simulación y en particular los de DS tienen la propiedad de generar datos simulados, a partir de la estructura modelada. Esto significa que, con relativamente pocos datos, se pueden construir modelos útiles, a partir de la premisa en la que se tenga un conocimiento suficiente sobre los niveles y flujos en el sistema real. Un ejemplo se muestra en el gráfico 5, en el que para la generación del modelo fueron necesarios apenas tres datos históricos sobre el porcentaje de adultos con obesidad en Estados Unidos.

De todas formas, los datos históricos no son las únicas fuentes de información para alimentar un modelo de DS. Son necesarios también conocimientos provenientes de la literatura científica, técnica y en muchos casos de levantamiento de campo, por medio de entrevistas con expertos, por ejemplo. En este sentido, a pesar de que en América Latina no existen muchas fuentes de datos históricos y mucho menos en formato de series temporales, se pueden explotar las otras fuentes de información.

4 Ver: https://www.iseesystems.com/.
5 Ver: https://www.ventanasystems.com/software/.

Gráfico 5. Ejemplo de datos necesarios para un modelo de dinámica de sistemas



Nota: Los puntos representan los datos históricos y la línea continua, la simulación.

Fuente: Essien *et al.*(2006)

Finalmente, las fuentes de conocimiento sobre la aplicación y el desarrollo de la DS son también importantes. Existen varios repositorios internacionales que sirven para este propósito, que ofrecen materiales didácticos para el autoaprendizaje, cursos de capacitación, videoaulas y otros, incluyendo principalmente los recursos educativos en el sitio web de la System Dynamics Society (SDS).[6]

Bibliografía

Barlas, Y. (1996). "Formal Aspects of Model Validity and Validation in System Dynamics". *System Dynamics Review*, vol. 12, nº 3, pp. 183-210.

Barra Montevechi, J. A.; de Carvalho Miranda, R.; de Queiroz, J. A.; dos Santos, C. H. y Leal, F. (2022). "Decision support in productive processes through DES and ABS in the Digital Twin era: a systematic

6 Ver: www.systemdynamics.org.

literature review". *International Journal of Production Research*, vol. 60, nº 8, pp. 2662-2681. DOI: 10.1080/00207543.2021.1898691.

Belhadi, A.; Kamble, S.; Maheshwari, P.; Mani, V. y Pundir, A. (2022). "Digital twin implementation for performance improvement in process industries. A case study of food processing company". *International Journal of Production Research*, pp. 1-23. DOI: 10.1080/00207543.2022.2104181.

Bendoly, E.; Linderman, K.; Oliva, R. y Sterman, J. D. (2015). "System dynamics perspectives and modeling opportunities for research in operations management". *Journal of Operations Management*, vol. 39-40, nº 1-5. DOI: http://dx.doi.org/10.1016/j.jom.2015.07.001.

Borshchev, A. (2013). *The Big Book of Simulation Modeling. Multimethod Modeling with AnyLogic 6*. USA: AnyLogic North America.

Cole, H. S. D.; Freeman, C.; Jahoda, M. y Pavitt, K. L. (1973). *Thinking about the Future: A Critique of the Limits to Growth*. Londres: Chatto & Windus.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (29 de septiembre de 2017). "Pós-graduação brasileira teve avanço qualitativo na última década". *Gov.br - Ministério da Educação*. Disponible en: https://www.gov.br/capes/pt-br/assuntos/noticias/pos-graduacao-brasileira-teve-avanco-qualitativo-na-ultima-decada.

Cosenz, F.; Dyner, I. y Herrera, M. M. (2019). "Assessing the effect of transmission constraints on wind power expansion in northeast Brazil". *Utilities Policy*, vol. 59. DOI: https://doi.org/10.1016/j.jup.2019.05.010.

Desai, A.; Ghaffarzadegan, N. y Hawley, J. (2014). "Research workforce diversity: The case of balancing national versus international postdocs in US biomedical research". *Systems research and behavioral science*, vol. 31, nº 2, pp. 301-315.

Diehl, E. W. y Sterman, J. D. (1995). "Effects of Feedback Complexity on Dynamic Decision Making". *Organizational Behavior and Human Decision Processes*, vol. 62, nº 2, pp. 198-215.

Downs, D.; Lindsay, A. y Lunn, K. (2003). "Business processes attempts to find a definition". *Information and Software Technology*, vol. 45, n° 15, pp. 1015-1019.

Duggan, J. (2016). *System dynamics modeling with R*. Suiza: Springer.

Dumas, M.; La Rosa, M.; Mendling, J. y Reijers, H. A. (2013). *Fundamentals of business process management*. New York: Springer.

Dyner, I. y Zuluaga, M. M. (2007). "Incentives for renewable energy in reformed Latin-American electricity markets: the Colombian case". *Journal of Cleaner Production,* vol. 15, n° 2, pp. 153-162.

Dyner, I.; Franco, C. J. y Jimenez, M. (2016). "Diffusion of renewable energy technologies: The need for policy in Colombia". *Energy*, vol. 111, pp. 818-829. DOI: https://dx.doi.org/10.1016/j.energy.2016.06.051.

Ekins, P. (1993). "'Limits to growth' and 'sustainable development': grappling with ecological realities". *Ecological Economics*, vol. 8, n° 3, pp. 269-288. DOI: https://doi.org/10.1016/0921-8009(93)90062-B.

Essien, J. D.; Homer, J. B.; Jones, A. P.; Milstein, B.; Murphy, D. L. y Seville, D. A. (2006). "Understanding diabetes population dynamics through simulation modeling and experimentation". *American Journal of Public Health*, vol. 96, n° 3, pp. 488-494.

Fiddaman, T. S. (2002). "Exploring Policy Options with a Behavioral Climate - Economy Model". *System Dynamics Review*, vol. 18, n° 2, pp. 243-267.

Ford, A. (2009). *Modeling the Environment*. Washington, DC: Island.

Forrester, J. W. (1958). "Industrial Dynamics: A Major Breakthrough for Decision Makers". *Harvard Business Review*, vol. 26, n° 4, pp. 37-66.

_____ (1961). *Industrial Dynamics*. Cambridge, MA: Productivity.

_____ (1973). *World Dynamics*. Cambridge, MA: Productivity.

_____ (1989). *The Beginning of System Dynamics*. Cambridge, MA: MIT System Dynamics Group.

_____ (1992). "Policies, Decisions, and Information Sources for Modeling". *European Journal of Operations Research*, vol. 59, n° 1, pp. 42-63.

Forrester, J. W. y Senge, P. M. (1980). "Tests for building confidence in system dynamics models". *System dynamics, TIMS studies in management sciences*, vol. 14, pp. 209-228.

Ghaffarzadegan, N.; Lyneis, J. y Richardson, G. P. (2011). "How small system dynamics models can help the public policy process". *System Dynamics Review*, vol. 27, nº 1, pp. 22-44. DOI: 10.1002/sdr.442.

Godinho Filho, M. y Uzsoy, R. (2009). "Efeito da redução do tamanho de lote e de programas de Melhoria Contínua no Estoque em Processo (WIP) e na Utilização: estudo utilizando uma abordagem híbrida System Dynamics - Factory Physics". *Production*, vol. 19, pp. 214-229.

_____ (2014). "Assessing the impact of alternative continuous improvement programmes in a flow shop using system dynamics". *International Journal of Production Research*, vol. 52, nº 10, pp. 3014-3031. DOI: 10.1080/00207543.2013.860249.

Grobbelaar, S. S. y Uriona-Maldonado, M. (2018). "Innovation System Policy Analysis through System Dynamics Modelling: A Systematic Review". *Science and Public Policy*, vol. 46, nº 1, pp. 28-44. DOI: 10.1093/scipol/scy034.

Hamacher, T. y Kotthoff, F. (2022). "Calibrating Agent-Based Models of Innovation Diffusion with Gradients". *Journal of Artificial Societies and Social Simulation*, vol. 25, nº 3. DOI: 10.18564/jasss.4861.

Houghton, J.; Martin-Martinez, E.; Samsó, R. y Solé, J. (2022). "PySD: System Dynamics Modeling in Python". *Journal of Open Source Software*, vol. 7, nº 78.

Landini, F.; Lee, K. y Malerba, F. (2017). "A history-friendly model of the successive changes in industrial leadership and the catch-up by latecomers". *Research Policy*, vol. 46, nº 2, pp. 431-446. DOI: http://doi.org/10.1016/j.respol.2016.09.005.

Meadows, D.; Meadows, D. H. y Randers, J. (2004). *Limits to growth: the 30-year update*. White River Junction: Chelsea Green.

Nelson, R. R. y Winter, S. (1982). *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press.

Niosi, J. (2004). "National systems of innovation are evolving complex systems". Presentado en el *13th International Conference on Management of Technology – IAMOT*. Washington, DC.

_____ (2010). *Building National and Regional Innovation Systems: Institutions for Economic Development*. Cheltenham, UK/Northampton, MA: Edward Elgar.

Ogata, K. (1998). *System Dynamics*. Upper Saddle River, NJ: Prentice Hall.

Pidd, M. (1998). *Computer Simulation in Management Science*. New York: John Wiley & Sons.

Rahmandad, H. y Sterman, J. (2008). "Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models". *Management Science*, vol. 54, nº 5, pp. 998-1014.

Raven, R. y Walrave, B. (2016). "Modelling the dynamics of technological innovation systems". *Research Policy*, vol. 45, nº 9, pp. 1833-1844. DOI: 10.1016/j.respol.2016.05.011.

Richardson, G. (1999). "System Dynamics". En Gass, S. y Harris, C. (eds.), *Encyclopedia of Operations Research and Management Science*. Amsterdam: Kluwer Academic.

Robinson, S. y Tako, A. A. (2010). "Model development in discrete-event simulation and system dynamics: An empirical study of expert modellers". *European Journal of Operational Research*, vol. 207, nº 2, pp. 784-794.

Seong, B. S. y Qudrat-Ullah, H. (2010). "How to do structural validity of a system dynamics type simulation model: The case of an energy policy model". *Energy Policy*, vol. 38, nº 5, pp. 2216-2224. DOI: https://doi.org/10.1016/j.enpol.2009.12.009.

Sterman, J. D. (1989). "Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment". *Management Science*, vol. 35, nº 3, pp. 321-339.

_____ (2000). *Business Dynamics. Systems Thinking and Modeling for a complex world*. Boston: Mc Graw Hill Higher Education.

_____ (2002). "All Models are Wrong: Reflections on Becoming a Systems Scientist". *System Dynamics Review*, vol. 18, nº 4, pp. 501-531.

_____ (2006). "Learning from evidence in a complex world". *American Journal of Public Health*, vol. 96, nº 3, pp. 505-514. DOI: http://dx.doi.org/10.2105/AJPH.2005.066043.

Vignieri, V. (2021). "Crowdsourcing as a mode of open innovation: Exploring drivers of success of a multisided platform through system dynamics modelling". *Systems Research and Behavioral Science*, vol. 38, nº 1, pp. 108-124. DOI: https://doi.org/10.1002/sres.2636.

Weinhardt, J. M., Hendijani, R., Harman, J. L., Steel, P., & Gonzalez, C. (2015). How analytic reasoning style and global thinking relate to understanding stocks and flows. *Journal of Operations Management*, 39, 23-30.

Capítulo 5
El método de revisión de la literatura estructurada para los estudios de CTI

Caroline Rodrigues Vaz, Mauricio Uriona-Maldonado

Introducción

La premisa de que la actividad científica puede rescatarse, estudiarse y sintetizarse a partir de la literatura publicada refuerza la necesidad de contar con métodos estructurados que ayuden a identificar las cuestiones de investigación pertinentes y a encontrar formas de estudiarlas. El método de revisión estructurada de la literatura sirve a este propósito al integrar las técnicas con un enfoque cuantitativo (análisis científico) y un enfoque cualitativo (análisis de contenido).

Mediante la bibliometría y la cienciometría es posible construir indicadores para evaluar la producción científica de las personas, las áreas de conocimiento y los países. Reunidos bajo la égida de los estudios métricos de la información, estos indicadores se han utilizado ampliamente en la evaluación de investigadores y áreas de conocimiento. Sin embargo, la evaluación de las investigaciones realizadas exclusivamente mediante análisis bibliométricos y cienciométricos es objeto de críticas, dado el carácter cuantitativo de estos enfoques. Además de provocar preguntas, ya que se refieren a la evaluación de la ciencia en sí misma y a la actividad científica realizada por los investigadores, la producción e interpretación de los indicadores bibliométricos es una tarea compleja que requiere que quienes los producen dominen conocimientos de diferentes áreas, como la ciencia de la información y la sociología de la ciencia, entre otras (da Silva, Massao Hayashi y Piumbato Innocentini Hayashi, 2011).

Al integrar estas dos visiones (cantidad y calidad), el examen estructurado de la literatura permite identificar la dinámica de crecimiento en los campos científicos (o áreas de investigación), así como las lagunas para fortalecer la creación de nuevas investigaciones (como las tesis doctorales, por ejemplo). Así, este capítulo tiene como objetivo presentar el método de revisión estructurada de la literatura y demostrar su aplicación en el área de la CTI y la sociedad, basado en el modelo SYSMAP (Scientometric and Systematic Yielding Mapping Process) (Rodrigues Vaz y Uriona-Maldonado, 2017).

Descripción del método SYSMAP

La revisión estructurada de la literatura se basa en el análisis científico y el análisis de contenido (o sistemático). El análisis científico ayuda a cuantificar el crecimiento de las ciencias sociales y las humanidades basado en el análisis de las citas. Creado en la década de 1950 por Eugene Garfield (2006), el método cienciométrico utiliza indicadores como el índice H para los autores, el factor de impacto y el indicador SCImago Journal Rank (SJR) para las revistas, y técnicas como el análisis factorial, de redes y de conglomerados.

Para este trabajo se considera como bibliometría "la aplicación de las matemáticas y los métodos estadísticos a los libros y otros medios de comunicación" (Pritchard, 1969: 349) y como cienciometría "el conjunto de métodos cuantitativos de investigación sobre el desarrollo de la ciencia como proceso de información" (Mulchenko y Nalimov, 1971).

De este modo, el análisis bibliométrico es útil para descifrar y cartografiar el conocimiento científico acumulado y los matices en evolución de campos bien establecidos, dando sentido a grandes volúmenes de datos no estructurados de forma rigurosa. Por lo tanto, los estudios bibliométricos bien realizados pueden sentar bases firmes para hacer avanzar un campo de forma nueva y significativa: permiten y facultan a los académicos para 1) obtener una visión general única, 2) identificar las lagunas de conocimiento, 3) derivar nuevas ideas para la investigación y 4) posicionar sus contribuciones previstas para el campo (Donthu *et al.*, 2021).

Por otra parte, el análisis de contenido, también conocido como revisión sistemática de la literatura, es un método cualitativo que tiene por objeto sintetizar los resultados de la investigación a partir del análisis

de grandes volúmenes de estudios científicos, sobre la base de los principios de transparencia, reproducibilidad y objetividad (Denyer, Smart y Tranfield, 2003; Colicchia y Strozzi, 2012).

Las limitaciones de uno se presentan como las ventajas del otro. Aunque la cienciometría es capaz de identificar macrorrelaciones cuantitativas entre autores, estudios o revistas, falla en los microanálisis individuales. El análisis de contenido ya es de por sí poderoso para identificar las características individuales de cada estudio, pero no tiene mecanismos para el macroanálisis. Así pues, la revisión estructurada de la literatura integra los puntos fuertes de ambos métodos.

Existen varios modelos propuestos por investigadores de diferentes áreas del conocimiento (ver el cuadro 1) que presentan procedimientos para realizar un análisis bibliométrico y un análisis sistemático.[1]

1 Por ejemplo: Antes *et al.* (2003), Colford Jr. *et al.* (2003), Burgess, Koroglu y Singh (2006), Faria Fernandes y Godinho Filho (2007), Godinho Filho y Lage Junior (2008), Cooper, Hedges y Valentine (2009), Chittó Stumpf y de Souza Vanz (2010), Pacheco Lacerda y Teixeira (2010), Bornia, Tezza y Vey (2010), Barratt, Choi y Li (2011), da Silva, Massao Hayashi y Piumbato Innocentini Hayashi (2011), da Rosa, Ensslin y Rolim Ensslin (2011), de Mesquita y do Rego (2011), Colicchia y Strozzi (2012), de Oliveira Lacerda, Ensslin y Rolim Ensslin (2012), Gohr *et al.* (2013), Dresch, Pacheco Lacerda y Valle Júnior Antunes (2015), Aisenberg Ferenhof y Fernandes (2016), Gough, Oliver y Thomas (2019), Donthu *et al.* (2021), Egger, Higgins y Smith (2022).

Cuadro 1. Características de los modelos publicados

Autores	Método	Características del modelo propuesto
Egger, Higgins y Smith (2022)	Análisis sistemático	1° Fuente y búsqueda 2° Selección de estudios 3° Evaluación de la calidad 4° Presentación de resultados e implicaciones
Donthu *et al.* (2021)	Análisis bibliométrico	1° Análisis del rendimiento 2° Cartografía científica 3° Evaluación del análisis bibliométrico
De Luca Canto (2020)	Análisis sistemático	1° Planificación 2° Definición de equipo 3° Elección del tema 4° Buscar una revisión sistemática sobre el tema elegido 5° Elaboración de la pregunta de investigación 6° Recopilación de información en el protocolo de investigación 7° Riesgo de sesgo 8° Metaanálisis 9° Sistema GRADE
Gough, Oliver y Thomas (2019)	Análisis sistemático	1° Introducción 2° Pregunta de revisión y metodología 3° Estrategia de búsqueda 4° Descripción y análisis de los estudios 5° Evaluación de la calidad y la pertinencia 6° Síntesis 7° Presentación

Fuente: elaboración propia

El modelo propuesto es el SYSMAP (Scientometric and Systematic Yielding Mapping Process), que tiene por objeto presentar de forma estructurada los principales procesos para llevar a cabo un examen de la literatura sobre un tema del que el investigador no tiene conocimiento o en el que el investigador trata de identificar detalles específicos sobre un aspecto y/o contexto particular, mediante la combinación de análisis científico y análisis de contenido.

El modelo SYSMAP consta de cuatro fases (como se muestra en la figura 1), que son: 1) construcción de la colección de artículos (muestra I); 2) proceso de filtrado; 3) análisis científico y 4) análisis de contenido (muestra II) y construcción de lagunas/oportunidades de investigación.

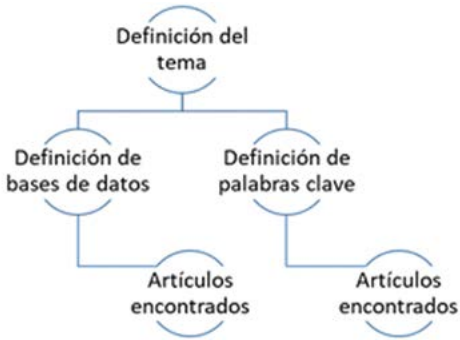
Figura 1. Modelo SYSMAP para la revisión estructurada de la literatura

Fuente: Vaz and Uriona-Maldonado (2017)

Fase 1: construcción de la colección de artículos (muestra I)

La construcción de la colección de artículos se divide en dos etapas, como se muestra en la figura 2.

Figura 2. Construcción de la colección de artículos de la fase 1



Fuente: elaboración propia (Rodrigues Vaz y Uriona-Maldonado, 2017)

Definición del tema

Para la correcta elaboración de la pregunta de investigación y su correcta estrategia de búsqueda, es necesario el uso de siglas, que son acrónimos que forman una palabra, utilizados como una estrategia didáctica para facilitar la elaboración de una buena pregunta de investigación.

De Luca Canto (2020) afirma que este tipo de estrategia didáctica sirve para elaborar una pregunta inicial en cualquier tipo de investigación, no solo en una revisión sistemática. El objetivo de esta estrategia didáctica es que el investigador no olvide ningún elemento esencial de la pregunta.

Según Oliveira Araújo (2020), existen los acrónimos: PICO (*population, intervention, comparison, outcome*), PICoS (*population, intervention, comparison, outcome, study type*), PICOT (*population, intervention, comparison, outcome, time*), PICOD (*population, intervention, comparison, outcome, design*), SPICE (*setting, perspective, intervention, comparison, evaluation*), SPIDER (*sample, phenomenon of interest, design, evaluation, research type*), PCC (*population, concept, context*), ECLIPSE (*expectation, client group, location, impact, professionals, service*) e TQO (*theme, qualifier, object of study*).

Para el SYSMAP se utilizará siempre el acrónimo TQO, ya que es una estrategia de búsqueda multidisciplinar, utilizada al inicio de una investigación para construir conocimiento en el investigador de forma general, de fácil construcción y enfoque simplificado, utilizada en cualquier área de conocimiento.

La estrategia TQO se estructura a partir de tres categorías: el tema, representado por el tema principal de la investigación, el calificador, representado por las características o situaciones relacionadas con el tema u objeto de investigación y, por último, el objeto, representado por un individuo, población, institución, dispositivo, procedimiento, etc. (Oliveira Araújo, 2020).

Por lo tanto, esta estrategia de búsqueda forma una pregunta de investigación, que debe ser clara, centrada y bien diseñada, es decir, que se pueda responder y sea innovadora. De Luca Canto (2020) explica que la mejor pregunta de investigación es la que interesa a un gran número de personas. No sirve de nada redactar una pregunta de investigación cuya respuesta solo interesa al primer autor o al coordinador del proyecto.

Oliveira Araújo subraya que esta estrategia no está pensada para la recuperación de pruebas. Su función principal es orientar al investigador hacia su tema de investigación, permitiendo el desarrollo de revisiones narrativas, estudios bibliográficos y análisis del estado del arte.

Así, el desarrollo de la estrategia debe responder las siguientes preguntas:

- T: ¿cuál es el tema principal a investigar?
- Q: ¿qué detalles específicos, o características, o factores culturales, o ubicación geográfica, o cuestiones de género, o cuestiones de raza, o procedimientos, etc., están relacionados con el objeto o el tema?
- O: ¿quién es el individuo, o la población, o la institución, o el dispositivo, etc. de la investigación?

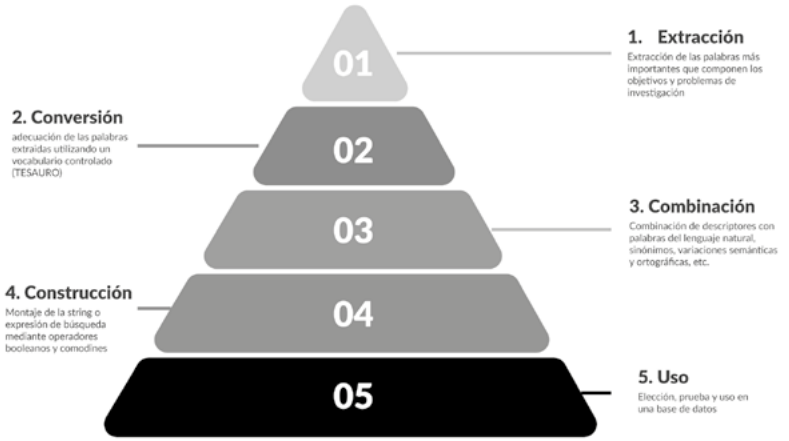
Por ejemplo:

- T: biogás
- Q: sistema tecnológico de innovación
- O: Brasil
- Pregunta de investigación: ¿cómo está el desarrollo del sistema tecnológico de innovación de biogás en Brasil?

Así, su uso pretende convertir el problema u objetivos de investigación en una estrategia de búsqueda compuesta por descriptores, lenguaje natural compuesto por operadores booleanos y caracteres comodín.

Para la construcción de esta estrategia de búsqueda adecuada, es necesario que el investigador utilice un vocabulario controlado, es decir, un tesauro. La figura 3 muestra los pasos de la estrategia de búsqueda para el inicio de la construcción de la búsqueda.

Figura 3. Etapas de estrategias de búsqueda



Fuente: elaboración propia

Definición de palabras clave

Es necesario para una búsqueda más eficiente que se realice una combinación de al menos tres palabras clave, es decir, identificar tres sinónimos para el mismo término que se desea buscar.

Además, es necesario para una búsqueda más específica, para obtener más resultados, utilizar dos o más términos, por ejemplo, el investigador quiere realizar una investigación en el área de sistemas tecnológicos de innovación en el sector del biogás. Su primer término de búsqueda sería "sistemas tecnológicos de investigación", seguido de "biogás".

Las definiciones y combinaciones de palabras clave deben estar en inglés, ya que la mayoría de las encuestas y publicaciones están en inglés. Así, el ejemplo anterior se vería así: "technological innovation systems" y "biogas". En las definiciones de las palabras se debe tener en cuenta también las siglas utilizadas para este término.

Para asegurar una búsqueda completa sobre el tema en cuestión, se utilizan los símbolos de truncamiento, que tienen por objeto recuperar las variaciones del plural y la ortografía. Los símbolos son: * (de cero a infinitos caracteres), * (un carácter), $ (cero o un carácter). Por ejemplo:

Child* (child, children, childbirth); Cell$ (cell, cells, cello); Dosto?evsk (Dostoyevsky, dostoievsky).

Además, se puede usar operadores booleanos para combinar y/o unir temas:

- AND: resulta en artículos que contienen tanto un término como otro. Ejemplo: "teoría de la racionalidad limitada" AND "psicológica".
- OR: resulta en artículos que contienen un término u otro. Ejemplo: "gestión de la innovación" OR "capital humano".
- AND NOT: resulta en artículos que contienen el primer término y no el segundo. Ejemplo: "medición del rendimiento" AND NOT "industria".

Se recomienda que para validar la palabra clave definida se encuentren al menos tres artículos aleatorios en Google Scholar.

Así, para el ejemplo mencionado en la esfera del biogás, tendríamos la siguiente combinación de palabras clave y operadores booleanos: "biogás" OR "biomethane" OR "biogas plants" AND "TIS" OR "Transitions" OR "Technological innovation systems".

También existe una forma más avanzada de combinación de palabras clave en las bases de datos cuando el investigador conoce todas las expresiones del tema a investigar, como se muestra en la investigación realizada por Alves Anacleto *et al.* (2017) en el área de la energía renovable, que compara la literatura científica y las patentes existentes sobre las placas solares aéreas, como se muestra en la cuadro 2.

Cuadro 2. Combinaciones de palabras clave

Palabras claves	Fórmula de búsqueda
Palabras clave para la revisión de literatura	TS=("AirborneWind Energy") OR TS=("AirborneWind Power") OR TS=("High Altitude Wind Energy") OR TS=("High AltitudeWind Power") OR TS=("Kite wind generator$") OR TS=("kite wind energy") OR TS=("Crosswind kite$") OR TS=("AirborneWind Turbine$") OR TS=("Flying Electric Generator$") OR TS=("Kite power") OR TS=("Kite energy") OR TS=("Pumping kite$") OR TS=("Lighter-Than-AirWind Energy System$") OR TS=("Kite model$") OR TS=("tethered undersea kite$") OR TS=("Kite-BasedWind Energy") OR TS=("kite wind power") OR TS=("Kite-Powered System$") OR TS=("Kite towed ship") OR TS=("crosswind towing") OR SO=("AirborneWind Energy") OR TS=(Parawing AND energy) OR (TS=(Kite) AND TS=("ship propulsion")) OR (TS=("Wind Power") AND TS=("flying kite$")) OR (TS=("kite") AND TS=("tracking control")) OR (TS=("kite") AND TS=("flight control")) OR (TS=("Kite generator") NOT DO=("10.1007/s00145-015-9206-4")) OR (TS=("Towing kite$" AND "wind energy")) OR (TS=("Kite system$") AND TS=("Wind energy")) OR (TS=("Kite system$") AND TS=("Power Generating")) OR (TS=("Power Kite$") AND TS=("Wind Energy")) OR (TS=("Tethered Airfoil$") AND TS=("Wind Energy")) OR (TS=("kite system") AND TS=(wind) NOT DO=(10.1007/BF00123534)) OR (TS=(kite) AND AU=("Creighton, Robert")) OR (TS=(Laddermill) AND TS=(kite)) OR DO=("10.2514/3.48003") OR DO=("10.1016/0167-6105(85)90015-7") OR DO=("10.1016/j.apenergy.2013.07.026") OR DO=("10.2514/1.31604") OR DO=("10.1002/rnc.1210")
Palabras clave para la revisión de patentes	TS=("AirborneWind Energy") OR TS=("AirborneWind Power") OR TS=("High Altitude Wind Energy") OR TS=("High AltitudeWind Power") OR TS=("Kite wind generator$") OR TS=("kite wind energy") OR TS=("Crosswind kite$") OR TS=("AirborneWind Turbine$") OR TS=("Flying Electric Generator$") OR TS=("Kite power") OR TS=("Kite energy") OR TS=("Pumping kite$") OR TS=("Lighter-Than-AirWind Energy System$") OR TS=("Kite model$") OR TS=("tethered undersea kite$") OR TS=("Kite-BasedWind Energy") OR TS=("kite wind power") OR TS=("Kite-Powered System$") OR TS=("Kite towed ship") OR TS=("crosswind towing") OR TS=("Parawing AND energy") TS=("AirborneWind Energy") OR TS=("AirborneWind Power") OR TS=("High Altitude Wind Energy") OR TS=("High AltitudeWind Power") OR TS=("Kite wind generator$") OR TS=("kite wind energy") OR TS=("Crosswind kite$") OR TS=("AirborneWind Turbine$") OR TS=("Flying Electric Generator$") OR TS=("Kite power") OR TS=("Kite energy") OR TS=("Pumping kite$") OR TS=("Lighter-Than-AirWind Energy System$") OR TS=("Kite model$") OR TS=("tethered undersea kite$") OR TS=("Kite-BasedWind Energy") OR TS=("kite wind power") OR TS=("Kite-Powered System$") OR TS=("Kite towed ship") OR TS=("crosswind towing") OR TS=("Parawing AND energy") OR TS=("Laddermill") OR TS=("flying kite$")

Fuente: Alves Anacleto *et al.* (2017)

Definición de la base de datos

Las bases de datos agrupan las revistas (científicas) según determinados criterios, principalmente de calidad: por ejemplo, los criterios de Web of Science para la indexación de revistas.

Por lo tanto, las bases de datos "venden" el acceso a los interesados de las revistas que contienen (por ejemplo: en Brasil la CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) compra la suscripción a bases de datos). Para garantizar un producto de calidad, las bases de datos tratan de seleccionar las "mejores" revistas o periódicos de cada área de conocimiento.

Por otro lado, para las revistas, el hecho de que "aparezcan" indexadas en una base de datos determinada, significa una mayor visibilidad, y por lo tanto un mejor posicionamiento frente a otras revistas "competidoras". También significa que se publicarán más y mejores artículos en ella, ya que la comunidad científica busca publicar su trabajo en las "mejores" revistas del área.

Esta fase depende del trabajo que el investigador quiera hacer, ya que es posible utilizar bases de datos como Web of Science, Scopus, ScienceDirect, PubMed, Emerald, SciELO, entre otras disponibles. Por ejemplo,

en Brasil estas bases están disponibles en el portal de CAPES (http://www.periodicos.capes.gov.br/). O también, realizar una búsqueda en revistas específicas, por ejemplo, un investigador desea realizar una búsqueda en el área de la economía de la innovación, utilizando solamente la revista *Research Policy*.

La base de datos más utilizada es Web of Science (base de datos especializada con unas 12.000 revistas indexadas, que ha venido recuperando trabajo desde 1990). Otra base que se ha utilizado desde su creación en 2004, es Scopus, desarrollada por Elsevier, ha estado recuperando trabajos desde 1996.

Además de las bases de datos científicas, las búsquedas deben realizarse en las bases grises, es decir, aquellas literaturas que no están indexadas en las bases principales. De Luca Canto (2020) señala que en estas bases se pueden encontrar investigaciones inéditas, disertaciones, tesis, documentos preimpresos o preliminares, documentos técnicos, informes estadísticos, memorandos, actas de conferencias, entre otros. Algunos ejemplos de estas bases grises son: Google Scholar, OpenGrey y ProQuest Dissertations & Theses. Por su parte, la gran ventaja de Google Scholar es que es gratuita, las otras dos requieren una suscripción especial para una mayor accesibilidad. Sin embargo, su fiabilidad y calidad de los datos son mejores que las de Google Scholar, puesto que esta no tiene una herramienta que identifique múltiples versiones del mismo documento y hay una limitación de mil accesos por resultado de la consulta.

Google Scholar utiliza la tecnología de Google para buscar no solo en las bases de datos científicas los trabajos académicos pertinentes, sino también en la web (por ejemplo, en el sitio web de una universidad o un congreso). A menudo es más eficiente para la búsqueda que las bases de datos tradicionales, pero hay reparos en su uso, especialmente para el análisis científico.

De Luca Canto explica que la búsqueda de literatura gris es una parte que no se puede olvidar o descuidar, porque identifica investigaciones que aún no han sido publicadas, convirtiéndose en algo esencial durante la búsqueda bibliográfica sistemática. De este modo, sirve para evitar el sesgo de publicación, ya que algunos estudios pueden no haber sido publicados en estas bases de datos o haber quedado solo en tesis y disertaciones.

Otro paso importante es la búsqueda de literatura adicional, es decir, la lectura de las listas de referencias de los artículos incluidos en busca de algún artículo que pueda haberse perdido en los procesos anteriores.

Todos estos procedimientos tienen por objeto garantizar que no se pierda ningún artículo durante el proceso de identificación.

En la misma línea, Dresch, Pacheco Lacerda y Valle Júnior Antunes (2015) destacan la "técnica de la bola de nieve", es decir, el uso de la ayuda de expertos, que consiste en presentar la lista de fuentes previamente preparada a un experto, solicitando que sugiera nuevas fuentes e indique otros expertos a los que consultar.

Finalmente, el investigador puede complementar su investigación cuando sea necesario con el proceso de búsqueda manual, en revistas, libros o cualquier otra fuente impresa especializada en el tema a investigar.

Tras definir las palabras clave y las bases de datos que utilizará el investigador, se lleva a cabo la búsqueda, en la que se identifica una cantidad equis de artículos, y se da al investigador con el nombre de "artículos encontrados" (ver la figura 2).

Fase 2: filtrado para obtener la muestra de artículos

En esta etapa es necesario cortar/filtrar los "artículos encontrados", porque muchas veces, cuando se utiliza más de una base de datos, se pueden encontrar artículos duplicados, y dependiendo de los términos de búsqueda definidos en el momento de la inserción en la base de datos junto con las palabras clave definidas, el artículo recuperado puede no ajustarse exactamente con lo que el investigador está buscando.

Así pues, esta fase consta de cinco pasos (figura 3) para realizar el filtrado de estos "artículos encontrados" que formarán los "artículos de la muestra I":

1. Verificación del área de búsqueda: al realizar una búsqueda en la base de datos, junto con las palabras clave, delimitar el área y subárea de búsqueda y el tipo de documento que se desea recuperar (artículos, libros, documentos de conferencias, entre otros que se encuentran disponibles en las bases de datos).
2. Identificación de los artículos duplicados: verificar que los artículos no se repitan.
3. Verificación del encuadre/alineación de los artículos por los títulos y resúmenes: se deben descartar los títulos y resúmenes de los

artículos encontrados que no se adhieran al trabajo que el investigador quiere hacer.

4. Identificación de la cantidad de citación del artículo: este paso solo se realiza cuando la investigación debe ser muy específica. Las citas de los artículos se pueden encontrar en la base de datos, si se utiliza Web of Science/Scopus o directamente en Google Scholar, utilizando el título del artículo.
5. Identificación de los artículos por el factor de impacto de las revistas.

Otra forma de filtrado que puede realizarse es a través del Methodi Ordinatio de Kovaleski, Negri Pagani y Resende (2015), cuando el investigador ya conoce el tema y puede tomar algunas decisiones.

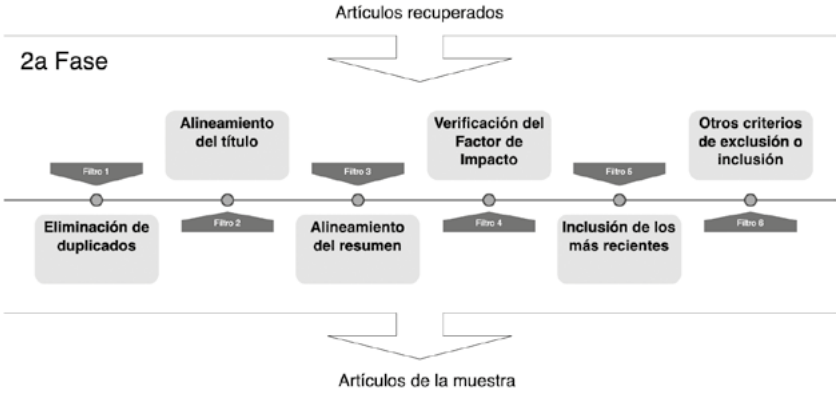
El Methodi Ordinatio utiliza una ecuación para clasificar los artículos, el Index Ordinatio (InOrdinatio), que tiene como objetivo seleccionar y clasificar los artículos según su relevancia científica, teniendo en cuenta los principales factores a considerar en un artículo científico: el factor de impacto de la revista en la que se publicó el artículo, el número de citas y el año de publicación (Kovaleski, Negri Pagani y Resende), como se muestra en la ecuación (1).

$$InOrdinatio = \frac{IF}{1000} + \alpha * [10 - (ResearchYear - PublishYear)] + (\sum C_i) \quad (1)$$

En esta etapa, el buscador puede incluir otros artículos que considere pertinentes en la búsqueda, que no se encontraron en la búsqueda, puesto que por alguna combinación de palabras clave ese artículo no se encontró.

Además, en esta etapa le corresponde al investigador hacer más filtros para delimitar los artículos que deben ser buscados (figura 4). Para el método SYSMAP el filtrado es solo de los artículos duplicados y alineados con el tema, porque el objetivo es que el investigador tenga un estado del arte para una mayor comprensión y entendimiento del tema.

Figura 4. Filtrado de la fase 2



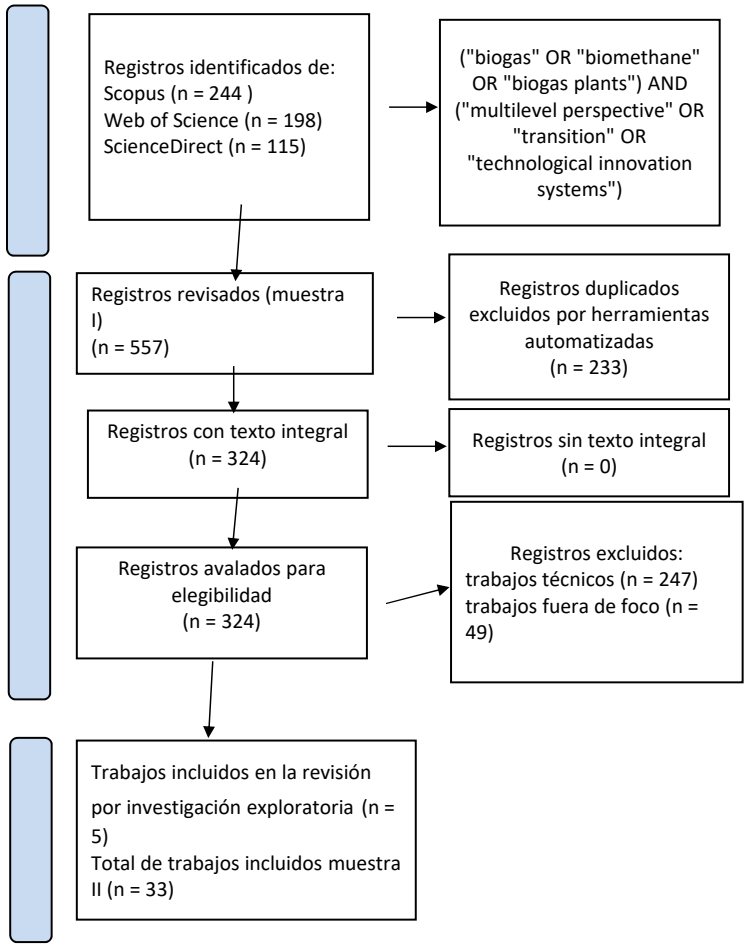
Fuente: elaboración propia (Rodrigues Vaz y Uriona-Maldonado, 2017)

Para el ejemplo de biogás utilizado en este trabajo, el diagrama 1 muestra los pasos de filtrado, la búsqueda se realizó en marzo de 2019, en las bases de datos Web of Science, Scopus y ScienceDirect, delimitando la búsqueda en el área de ingeniería, obteniendo 557 artículos en bruto, de los cuales 233 se repitieron. A continuación, se seleccionaron 78 artículos sobre la base de los títulos y resúmenes alineados con el tema. Posteriormente, se seleccionaron 28 artículos según el factor de impacto de la revista que lo incluyó y, finalmente, se incluyeron 5 artículos de investigación exploratoria, obteniendo "artículos de muestra I" de $n = 33$.

Para la representación gráfica de los pasos de la investigación, es necesario utilizar un protocolo. El protocolo es un proyecto, un plan, y debe contener toda la información necesaria para guiar el trabajo de los investigadores. Un ejemplo de protocolo ampliamente utilizado y recomendado en la investigación científica es el checklist PRISMA-P.

En SYSMAP se utilizará siempre como base el diagrama de flujo PRISMA-P, para registrar y presentar toda la información de los pasos de inclusión y exclusión con sus respectivas justificaciones de los trabajos a utilizar en la investigación, un ejemplo se puede ver en el diagrama 1.

Diagrama 1. Pasos de filtrado



Fuente: elaboración propia con base en Akl *et al*. (2021)

Riesgo de sesgo

Una cuestión importante a la hora de filtrar los artículos es el riesgo de sesgo que el investigador puede provocar en la investigación. Un sesgo de investigación es un error sistemático que afecta la validez del estudio, es

decir, el resultado de un error sistemático en el diseño, la realización o el análisis estadístico de un estudio, que puede provocar una distorsión en el resultado. Esto ocurre debido a fallas en el método, en la forma de recoger o evaluar la información de la investigación.

Algunos tipos de riesgo de sesgo que puede encontrar o realizar el investigador en los estudios científicos, debido a la asociación entre la exposición y el resultado:

- Sesgo de tiempo de publicación: publicación rápida o tardía de los resultados de la investigación, según la naturaleza y la dirección de los resultados.
- Sesgo de publicación múltiple: resultados de investigación publicados en varias fuentes, según la naturaleza y la dirección de los resultados, que pueden computarse más de una vez en el metaanálisis.
- Sesgo del lugar de publicación: publicación de los resultados en revistas de fácil acceso o niveles de indexación en bases de datos estándar, según la naturaleza y la dirección de los resultados.
- Sesgo de citación: la citación o no de los resultados de la investigación, en función de la naturaleza y la dirección de los resultados, con tendencia a una menor citación de los estudios con resultados negativos.
- Sesgo lingüístico: la publicación de los resultados de la investigación en un idioma específico, según la naturaleza y la dirección de los resultados o la exclusión de los estudios publicados en otro idioma en el proceso de selección de artículos.
- Sesgo de publicación selectiva de resultados: la comunicación de resultados para algunos resultados y no para otros, según la naturaleza y la dirección de los hallazgos.

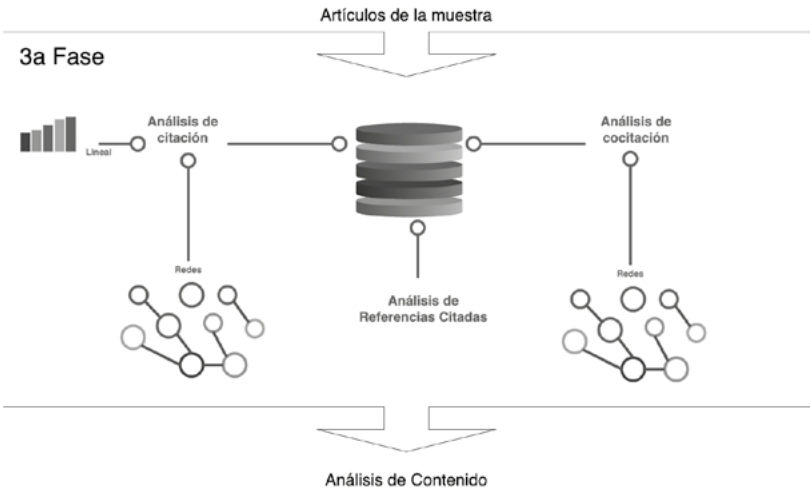
Así, la evaluación del riesgo de sesgo de los estudios contribuirá a juzgar la solidez de las pruebas de la revisión de la literatura. Si existe un sesgo metodológico en los estudios incluidos, el resultado de la revisión puede verse comprometido. Así, el análisis del riesgo de sesgo ayuda a interpretar si los datos obtenidos de los estudios incluidos son fiables para orientar las decisiones (De Luca Canto, Massignan y Stefani, 2021).

Existen numerosas herramientas publicadas en la literatura para este fin y cada diseño de estudio presenta una herramienta específica, algunos ejemplos son: RevMan (ReviewManager), el método del Joanna Briggs Institute, AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews)), entre otros.

Fase 3: análisis científico

El análisis científico es una técnica que permite hacer un mapa de los principales autores, revistas y palabras clave de un tema determinado. Macedo dos Santos, Silva Santos y Uriona-Maldonado (2010) afirman que estas técnicas son herramientas que se basan en una base metodológica teórica científicamente reconocida, que permite utilizar métodos estadísticos y matemáticos para cartografiar la información de los registros bibliográficos de los documentos almacenados en las bases de datos. Por lo tanto, esta fase se divide en tres etapas, como se muestra en la figura 5.

Figura 5. Análisis científico de la fase 3

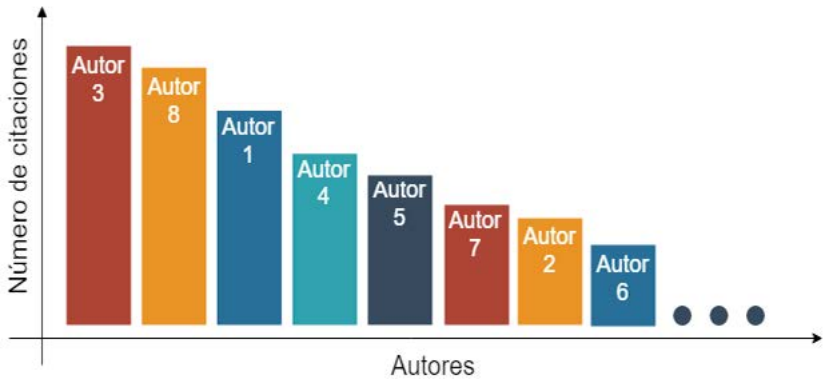


Fuente: elaboración propia (Rodrigues Vaz y Uriona-Maldonado, 2017)

Análisis de citas

La relevancia de cada "unidad de análisis" se identifica a partir de las veces que esta "unidad de análisis" ha sido citada por otros trabajos, como se muestra en el gráfico 1. La unidad de análisis está representada por: autores, revistas, documentos, países, instituciones educativas, periodicidad, palabras clave, entre otros.

Gráfico 1. Ejemplo de representación de autores más citados

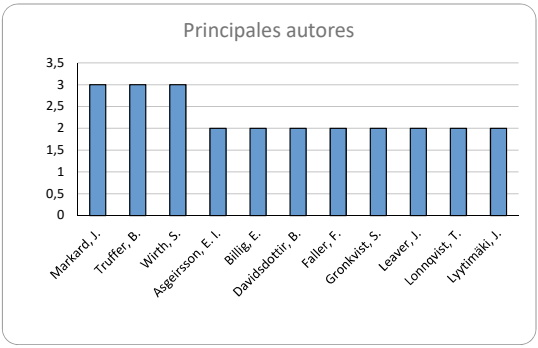


Fuente: elaboración propia

Citas de autores, revistas, países, instituciones educativas, palabras clave: los análisis se construyen por el número de veces que aparecen repetidos en la muestra de artículos. Por ejemplo, en la muestra de biogás para la cita de autor, se encontraron 113 autores y coautores, como puede verse en el gráfico 2.

Gráfico 2. Autores más relevantes

Markard, J.	3
Truffer, B.	3
Wirth, S.	3
Asgeirsson, E. I.	2
Billig, E.	2
Davidsdottir, B.	2
Faller, F.	2
Gronkvist, S.	2
Leaver, J.	2
Lonnqvist, T.	2
Lyytimäki, J.	2



Fuente: elaboración propia

Los autores J. Markard, B. Truffer y S. Wirth son los más influyentes de la muestra y han participado en tres publicaciones cada uno.

Frecuencia de las citas: el número de documentos (artículos) que se repiten por año. Por ejemplo, para la muestra de biogás, el gráfico 3 muestra la evolución de las publicaciones. El crecimiento del tema del biogás puede observarse en los estudios de los sistemas tecnológicos nacionales de innovación. Básicamente, las publicaciones han crecido de manera constante, mostrando un crecimiento exponencial hasta el año 2018.

Gráfico 3. Periodicidad de las publicaciones

2008	1
2009	2
2010	0
2011	0
2012	0
2013	2
2014	3
2015	9
2016	7
2017	7
2018	11
2019	2

Fuente: elaboración propia

Citas de documentos: los documentos (artículos) más relevantes serán aquellos con una cantidad de citas. Sin embargo, estos datos dependen de la base de datos que el investigador utilice para realizar su investigación, por ejemplo, la Web of Science aporta la cantidad de citas que el artículo adquiere con el tiempo. Otra herramienta que el buscador puede decidir utilizar es Google Scholar, actualmente existe un software que realiza este análisis de forma automática (Zotero), sin necesidad de utilizarlos buscadores que realizan la búsqueda de forma manual. Por ejemplo, en la investigación realizada por Rodrigues Vaz y Uriona-Maldonado (2015) en el ámbito de la transición sociotécnica y de sistemas de innovación tecnológica, los investigadores realizaron el análisis de citaciones de los documentos más relevantes de la muestra, separando por título, revista, año, cantidad de citas y área temática del artículo, como se muestra en el cuadro 3.

Cuadro 3. Los quince documentos más relevantes

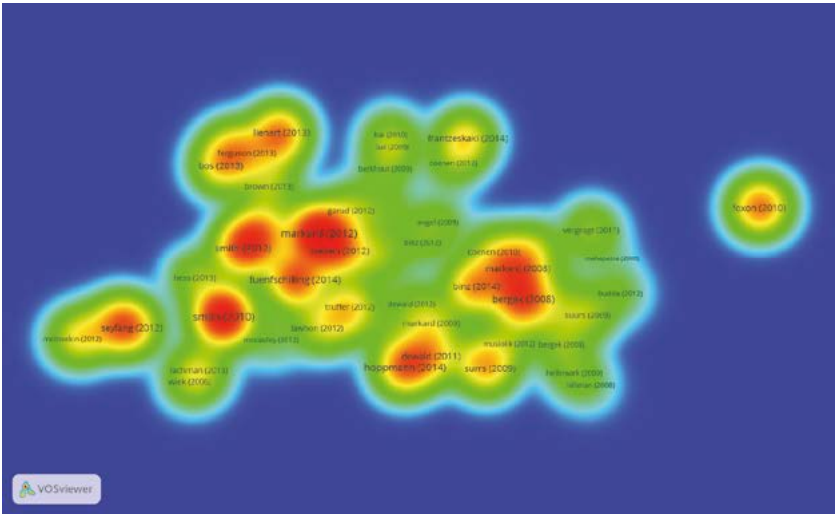
Orden	Autores	Título	Revista	Año	Citas	Tema
1	Markard, J.; Raven, R. y Truffer, B.	"Sustainability transitions: An emerging field of research and its prospects"	*Research Policy*	2012	96	ST
2	Bergek, A.; Carlsson, B.; Jacobsson, S.; Lindmark, S. y Rickne, A.	"Analyzing the functional dynamics of technological innovation systems: A scheme of analysis"	*Research Policy*	2008	96	TIS
3	Markard, J. y Truffer, B.	"Technological innovation systems and the multi-level perspective: Towards an integrated framework"	*Research Policy*	2008	86	MLP/ST/TIS
4	Benneworth, P.; Coenen, L. y Truffer, B.	"Toward a spatial perspective on sustainability transitions"	*Research Policy*	2012	55	ST
5	Raven, R. y Smith, A.	"What is protective space? Reconsidering niches in transitions to sustainability"	*Research Policy*	2012	53	SNM/ST
6	Smith, A. y Voss, J. P. & Grin, J.	"Innovation studies and sustainability transitions: The allure of the multi-level perspective and its challenges"	*Research Policy*	2010	52	ST/MLP
7	Coenen, L.; Farla, J.; Markard, J. y Raven, R.	"Sustainability transitions in the making: A closer look at actors, strategies and resources"	*Technological forecasting and social change*	2012	29	ST
8	Bergek, A.; Jacobsson, S. y Sanden, B. A.	"'Legitimation' and 'development of positive externalities': two key processes in the formation phase of technological innovation systems"	*Technology Analysis and Strategic Management*	2008	26	TIS
9	Coenen, L. y Truffer, B.	"Environmental Innovation and Sustainability Transitions in Regional Studies"	*Regional Studies*	2012	20	ST
10	Surrs, R. A. A & Hekkert, M. P.	"Cumulative causation in the formation of a technological innovation system: The case of biofuels in the Netherlands"	*Technological forecasting and social change*	2009	20	TIS
11	Musiolik, J..; Markard, J. y Hekkert M. P.	"Networks and network resources in technological innovation systems: Towards a conceptual framework for system building"	*Technological forecasting and social change*	2012	19	TIS
12	Coenen, L. y Truffer, B.	"Places and Spaces of Sustainability Transitions: Geographical Contributions to an Emerging Research and Policy Field"	*European Planning Studies*	2012	17	ST
13	Lawhon, M. y Murphy, J. T.	"Socio-technical regimes and sustainability transitions: Insights from political ecology"	*Progress in Human Geography*	2012	17	ST/MLP
14	Coenen, L. y Díaz López, F. J.	"Comparing systems approaches to innovation and technological change for sustainable and competitive economies"	*Journal of Cleaner Production*	2010	15	ST
15	Fuenfschilling, L. y Truffer, B.	"The structuration of socio-technical regimes-Conceptual foundations from institutional theory"	*Research Policy*	2014	14	MLP

Fuente: Rodrigues Vaz y Uriona-Maldonado (2015)

En este tipo de análisis el investigador puede combinar con otros enfoques de análisis para obtener mayores descubrimientos en el área,

como en el ejemplo que utilizan los autores Rodrigues Vaz y Uriona-Maldonado, quienes además de identificar los principales documentos con mayor número de citas, también identificaron las relaciones entre los documentos, a través del análisis de la colaboración en red de los autores (se tratará con mayor detalle en el próximo apartado), como se muestra en la figura 6, para tener una mayor comprensión de los datos y entendimiento del estado del arte.

Figura 6. Red de colaboración de los autores



Fuente: Rodrigues Vaz y Uriona-Maldonado (2015). Visualización de VOSviewer, 311 artículos, número mínimo de citas = 15,57 artículos y 13 grupos y 175 enlaces

La figura 6 está construida con el software VOSviewer, representada con el modo "vista de densidad", que permite la identificación de "temas candentes", por los cuales varios documentos se relacionan entre sí, formando "zonas de calor" en rojo.

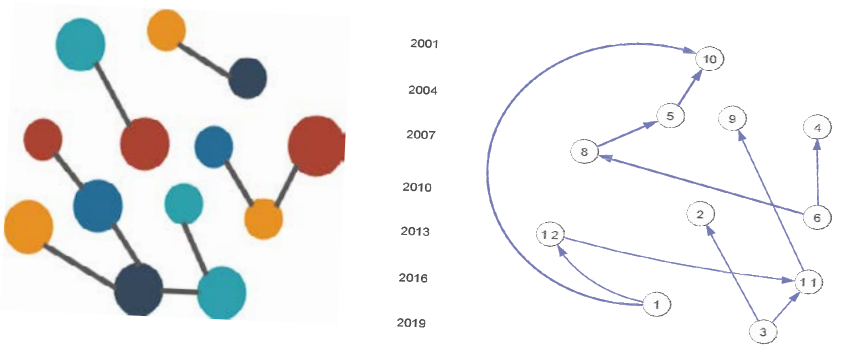
En el trabajo de Rodrigues Vaz y Uriona-Maldonado, la red de citas de documentos incluye todos los documentos con quince o más citas, totalizando 57 documentos, 13 agrupaciones y 175 enlaces. El gráfico se interpreta mediante la representación de colores y círculos, es decir, la mayor "zona de calor" está formada por las obras de Markard, Raven y Truffer (2012); Benneworth, Coenen y Truffer (2012); Raven y Smith (2012)

y Fuenfschilling y Truffer (2014), esta última centrada en el desarrollo de la teoría de los regímenes sociotécnicos. La segunda "zona de calor" más grande está representada por la labor de Markard y Truffer (2008), un examen de las transiciones hacia la sostenibilidad; Bergek *et al.*, 2008, sobre las funciones de los sistemas de innovación y el esquema de análisis; Binz, Coenen y Truffer (2014), sobre la necesidad de considerar diferentes escalas espaciales y perspectivas para el SIT, y Coenen y Díaz López (2010), que ofrece una revisión sistemática de la literatura sobre tres enfoques de sistemas en la literatura de las transiciones de sostenibilidad (sistemas de innovación sectorial, SIT y sistemas sociotécnicos).

Análisis por cocitación

La relevancia de cada "unidad de análisis" se identifica a partir de los momentos en que esta "unidad de análisis" aparece citada simultáneamente por otros trabajos, como se muestra en la figura 7.

Figuras 7. Ejemplos de representaciones de redes



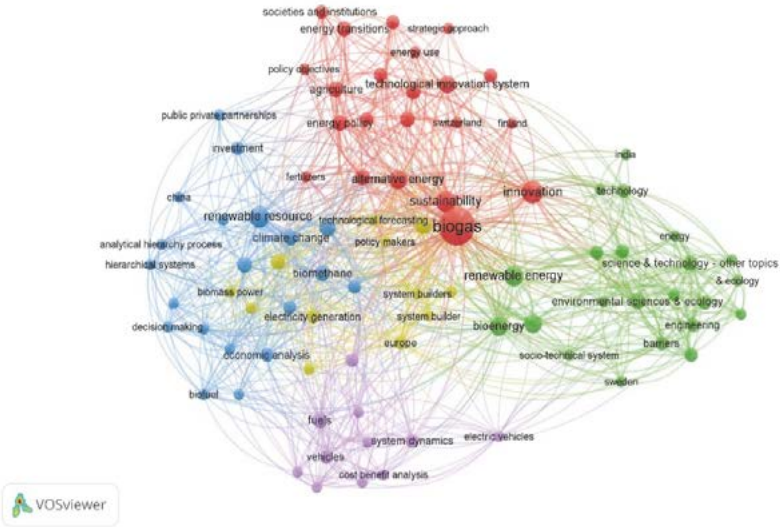
A. Red B. Red de colaboración

Fuente: elaboración propia

Para la cocitación de autores, revistas, documentos, palabras clave, países, instituciones educativas: los análisis de cocitación se construyen por el número de veces que aparecen interrelacionados, es decir, por la cooperación entre "unidad de análisis". Para este tipo de análisis es necesario utilizar un software específico para construir gráficos

bibliométricos. Por ejemplo, en la muestra de biogás, la figura 8 muestra la cocitación de palabras clave, con un total de 256, formadas por cinco cúmulos representados por colores, 2549 enlaces, construidos en el software VOSviewer. Las palabras clave se clasificaron por el número de ocurrencias, de modo que las más utilizadas aparecen en un tamaño más grande. Sin embargo, a menudo se utilizan palabras diferentes con significados similares (por ejemplo, energía renovable o energía alternativa), por lo que vale la pena recordar que el tratamiento de los datos contiene inferencias del investigador.

Figura 8. Cocitación de palabras clave



Fuente: elaboración propia

El cúmulo centralizado en el biogás (rojo) está conectado a diferentes nodos, como energía alternativa o renovable, que son obvios hasta cierto punto, pero los nodos más interesantes son los más pequeños (en dos conexiones): vehículos eléctricos (púrpura) y cooperativas. Estos dos nodos muestran que el biogás va más allá del autoconsumo del productor, para interactuar con nichos aparentemente no relacionados, como los coches eléctricos; sin embargo, el segundo nodo indica cómo el biogás requiere

cambios en los factores sociales para su difusión, las cooperativas son el ejemplo perfecto de esta necesidad.

El segundo grupo se centra en la lente teórica de la CTI (verde) con conexiones en el desarrollo tecnológico y los diferentes usos de la biomasa, como la gasificación. La gasificación es una técnica de producción de gas renovable diferente de la digestión anaeróbica (DA), aunque todavía no ha alcanzado un nivel tecnológico comercial y por lo tanto parece distante del biogás. El papel de Europa también puede reconocerse como un nudo importante, que representa el estado de la técnica en el desarrollo tecnológico del biogás y los enfoques analíticos relacionados con las transiciones, como los CTI.

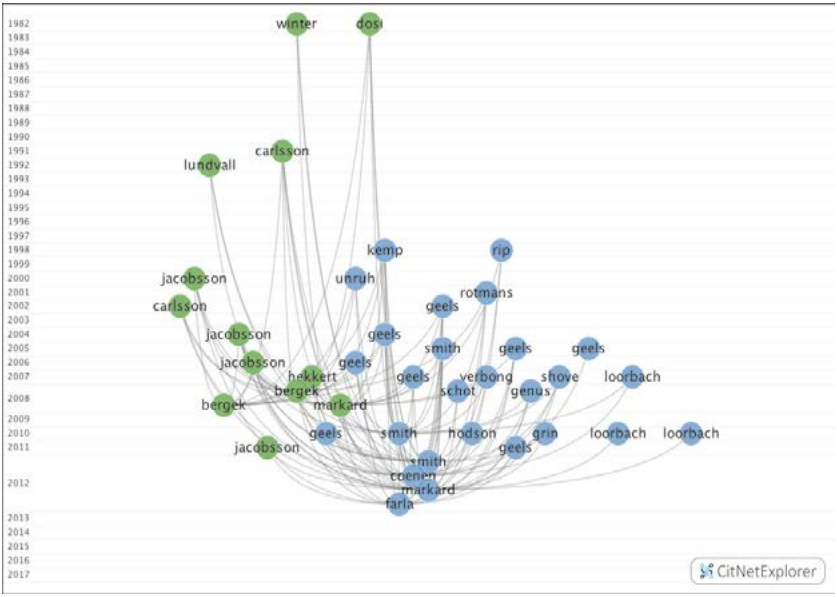
El tercer grupo es el de ingeniería (azul), vinculado a palabras clave como: innovación, ciencias ambientales, gestión, diseño industrial y aplicación. Este grupo representa el paradigma actual, las técnicas y prácticas que han guiado el comportamiento de búsqueda de los ingenieros en los últimos años.

El cuarto grupo (amarillo) destaca las palabras: sostenibilidad y bioeconomía, pero son otras dos más pequeñas las que llaman la atención: sistemas y dinámica. Estas ilustran la necesidad de modelar los sistemas de bioeconomía, ya que la comprensión de estos sistemas rara vez es lineal y suele manifestarse en el largo plazo, lo que dificulta la percepción de sus fallas y la toma de decisiones.

Los otros grupos son pequeños pero relevantes, por ejemplo, el biometano (cian) es una forma de comercializar el biogás que ya está ganando legitimidad para formar el propio grupo, especialmente en el caso de los países desarrollados.

Otro ejemplo de red, son las redes de colaboración por histograma, presentadas por primera vez por E. Garfield con el fin de aclarar las conexiones de red entre nodos (Garfield y Pudovkin, 2004; Garfield, 2009) a través del software CitNetExplorer. Estas redes permiten hacer un diagnóstico cronológico de las publicaciones más relevantes, pudiendo identificar las transiciones y evoluciones de los temas y áreas investigadas. Por ejemplo, en las investigaciones realizadas por Rodrigues Vaz y Uriona-Maldonado (2015) para el ámbito de la transición sociotécnica y el sistema de innovación tecnológica, los investigadores crearon el histograma de la evolución de este tema, como se puede ver en el gráfico 4.

Gráfico 4. Histograma



Fuente: Rodrigues Vaz y Uriona-Maldonado (2015)

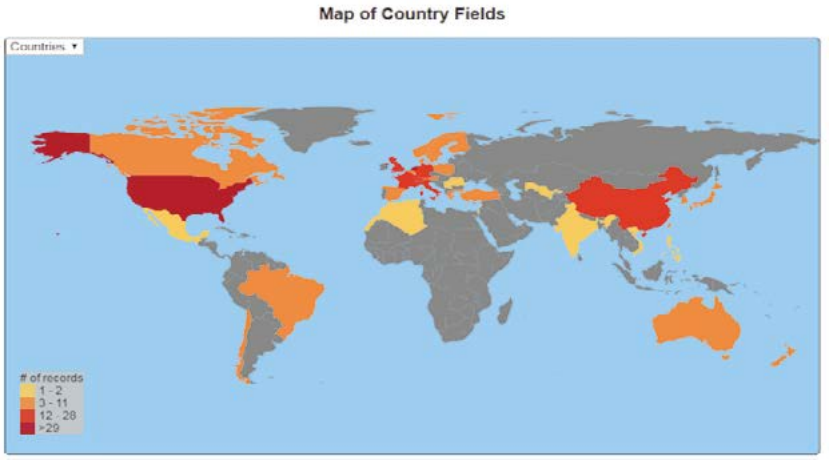
El histograma muestra dos grandes cúmulos (representados por los colores verde y azul) que evolucionaron a lo largo del tiempo, ambos, sin embargo, se desarrollaron bajo las contribuciones de Dosi (1982) y Nelson y Winter (1982), quienes desarrollaron la idea de regímenes tecnológicos y paradigmas/trayectorias tecnológicas, a pesar de ser fundamentales para proponer la perspectiva evolutiva del comportamiento de las empresas. Ambos grupos (verde y azul) fueron inspirados por ellos.

El grupo verde también recibió contribuciones teóricas de la obra de Carlsson y Stankiewicz (1991), con la primera aparición del concepto de sistemas de innovación tecnológica, y la obra de Lundvall (1992), con el primer libro sobre el sistema nacional de innovación. Otras contribuciones importantes al grupo ecológico fueron la labor de Jacobsson y Johnson (2000), que examinan cuestiones clave para la difusión de las tecnologías de energía renovable, y varias contribuciones más a la bibliografía sobre sistemas de innovación tecnológica (Bergek *et al.*, 2008). Otros trabajos que aparecen en el cúmulo verde son los ya mencionados en el análisis anterior (Hekkert *et al.*, 2007; Bergek *et al.*, 2008; Markard y

Truffer, 2008) y revisiones de la literatura de Markard, Raven y Truffer (2012) y Benneworth, Coenen y Truffer (2012).

Otra forma de red de colaboración es geográfica o espacial, este análisis tiene por objeto demostrar la densidad de las publicaciones por regiones, como se muestra en el mapa 1. Identificar las tendencias regionales sobre el tema que se estudia. En estos casos, hay un software específico para realizarlo. Por ejemplo, el mapa 1 muestra la investigación realizada por los autores de este artículo en 2017 con la combinación de palabras clave de coches eléctricos en la base de datos de Web of Science utilizando el software Vantage Point.

Mapa 1. Red de colaboración geográfica



Fuente: elaboración propia

En el mapa se puede observar que las densidades geográficas están representadas en 4 cúmulos (rojo, naranja oscuro, naranja claro y amarillo). La mayor densidad de publicación sobre el tema de los coches eléctricos se encuentra en Estados Unidos y el oeste de Canadá, seguido por China.

Este gráfico coincide con Røstvik (2018), quien explica que los carros eléctricos e híbridos ofrecen una alternativa prometedora para reducir las emisiones de carbono. Mientras que, en los países desarrollados como Estados Unidos, Noruega, Francia y Alemania, el parque automotor de China aumenta cada año, en los países en desarrollo su tasa de adopción

ha sido bastante baja. Especialmente, Noruega, por ejemplo, es un caso ejemplar de éxito en lo que se refiere a la difusión del automóvil eléctrico, motivada por generosos incentivos y apoyo político desde 1994.

Análisis de las referencias

Es sumamente importante realizar el análisis de las referencias de la muestra de artículos, para identificar los documentos más relevantes (los más repetidos en todos los documentos). A veces, debido a la combinación de palabras clave, definida por el investigador, no puede hallar ese artículo o trabajo que se considera esencial para el área que el investigador está investigando. Hay softwares que efectúan estos análisis, como HistCite, pero solo se pueden realizar con datos de la Web of Science. Por ejemplo, en la investigación sobre el biogás que presentamos, la muestra I dio lugar a 33 artículos, compuestos de 1320 referencias, las cuales pertenecían a libros, artículos de revistas, documentos de congresos, tesis, disertaciones y sitios especializados en la materia (como se muestra en el cuadro 4).

Cuadro 4. Lista de referencias de los artículos de la muestra I

Documentos	Cantidad
Libros	287
Artículos de revistas	678
Artículos de congresos	275
Tesis	13
Disertaciones	47
Sitios especializados	20

Fuente: elaboración propia

De las 1320 referencias analizadas, se pudo identificar que 435 documentos se repetían y que cuatro documentos (artículos de revistas) eran los propios artículos de muestreo.

Para este análisis de las referencias se puede llevar a cabo el análisis de las citas de autores, revistas, países, instituciones educativas, palabras clave, por el número de veces que aparecen repetidas en la muestra de artículos (solo que ahora para las referencias).

En este análisis de referencias, es importante llevar a cabo el análisis de las citas de los documentos, que tiene por objeto identificar los documentos más relevantes y esenciales para el tema investigado. Luego se incluyen estos documentos (preferentemente artículos de revistas) en la muestra de la cienciometría de la investigación para un mayor alcance del tema.

Después de todos estos análisis científicos de los artículos de la muestra, con el logro de la inclusión y exclusión, el investigador llega a la fase 4 de análisis de contenido, es decir, de lectura integral de los documentos e identificación de las oportunidades y lagunas de investigación en el área temática investigada.

Fase 4: análisis sistemático y/o análisis de contenido

Las revisiones sistemáticas permiten recuperar una gran cantidad de conocimientos acumulados en los estudios primarios y, al mismo tiempo, tienden a facilitar el desarrollo de investigaciones que aún necesitan una comprensión más profunda del tema abordado. Por lo tanto, el análisis debe ser algo más que una colección de los diferentes elementos investigados. Se espera que la consolidación y agregación de los resultados de los estudios primarios dé lugar a nuevos conocimientos (Dresch, Pacheco Lacerda y Valle Júnior Antunes, 2015).

Según de Oliveira Lacerda, Ensslin y Rolim Ensslin (2012), el análisis sistémico es el proceso utilizado para una visión conjunta del mundo (afiliación teórica) y explicado por su lente, analizando una muestra de elementos representativos de un determinado tema de búsqueda, buscando pruebas para cada lente y, a nivel global, para una perspectiva conjunta, reflexiones y oportunidades (o escasez) de conocimiento basado en muestras.

En la misma línea, Bardin (2011) explica que el análisis del contenido puede ser tanto cualitativo-cuantitativo (a menudo así es la información que surge delcarácter del contenido) como cualitativo (es la presencia o ausencia de un determinado carácter lo que se tiene en cuenta).

Dresch, Pacheco Lacerda y Valle Júnior Antunes explican que una revisión sistemática sigue un método explícito, planificado, responsable y justificable, al igual que los estudios primarios. Este método debe estar diseñado para garantizar que la revisión esté libre de sesgos, sea rigurosa, auditable, reproducible y actualizada.

En esta etapa, el modelo SYSMAP puede incluir artículos de otras fuentes y también puede considerarse como libros, monografías, tesis y disertaciones, formando así la muestra II, considerada en este trabajo, es decir, los documentos que serán analizados según su contenido para identificar la brecha de investigación sobre el tema definido.

Primero, para iniciar el análisis de contenido, es necesario construir una pregunta de búsqueda. Esta pregunta debe formularse de acuerdo con las necesidades de investigación del investigador, sobre la base del acrónimo utilizado para la estructuración de las palabras clave.

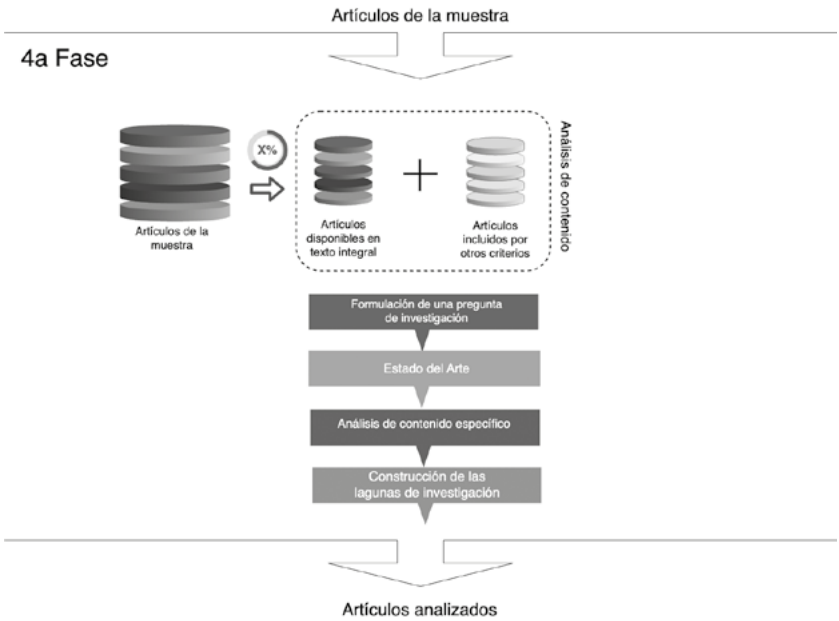
Por ejemplo, en la investigación sobre el biogás, la pregunta sería: "¿cómo se está desarrollando el sistema tecnológico nacional de innovación del biogás en Brasil?".

Los autores deben identificar y separar estos documentos en ocho categorías de análisis (figura 10):

1. Representatividad del artículo: los documentos deben separarse según el número de autores citados y por la cantidad de citas de cada artículo (por ejemplo, de los últimos tres años).
2. Tipo de investigación: se divide en teórica y empírica. Todavía se puede identificar cuál es el sector, región, de la investigación (estado, país, ciudad, entre otros) y cuál es el foco de la investigación, el proceso, el producto y/o el servicio de la organización.
3. Objetivo y problema de la investigación: identificar cuáles son los objetivos de la obra y lo que se quiere conseguir.
4. Conceptos: identificar qué definiciones han sido adoptadas por los autores sobre el tema a investigar y qué autores han sido citados.
5. Metodología: puede ser de tres maneras, cualitativa (trabajo de revisión de la literatura), cuantitativa (estudios que utilizan la simulación, estadísticas, encuestas, estudios de casos y multicasos, ecuaciones estructuradas, entre otros) y mixta (utilizando la combinación de lo cualitativo y lo cuantitativo).
6. Contribución principal: identificar las contribuciones/resultados que el trabajo ha proporcionado, tanto teóricas como prácticas, para la comunidad científica y las organizaciones empresariales.

7. Principales resultados: identificar qué resultados ha proporcionado el trabajo, tanto teóricos como prácticos, para la comunidad científica y las organizaciones empresariales.
8. Trabajos futuros: identificar si los autores de estos trabajos presentan recomendaciones de trabajos futuros para la continuación de las investigaciones sobre este tema.
9. Para una mejor comprensión de esta fase, el cuadro 5 trae el ejemplo de una hoja de cálculo de Excel con el tema de análisis del estado del arte del biogás. Hay programas de software específicos que realizan este tipo de análisis como el StArt.

Figura 10. Fase 4: análisis de contenido



Fuente: elaboración propia (Rodrigues Vaz y Uriona-Maldonado, 2017)

Cuadro 5. Hoja de cálculo Excel

Autor	Año	Título	Revista	¿Cuál era el problema y el objetivo del artículo?	¿Cuál es la definición utilizada por los autores?	¿Cuál es la metodología del artículo? ¿Teórica o empírica?	Si es empírica, ¿cuál es la ubicación, el sector, el proceso de aplicación de la metodología?	¿Cuáles son los principales resultados encontrados?	¿Cuáles son las recomendaciones para el trabajo futuro?
Al-Addous, M.; Alnaief, M.; Bdour, M. y Saidan, M. N.	2018	"Evaluación de la producción de biogás a partir de la codigestión de desechos de alimentos municipales y lodos de aguas residuales en los campos de refugiados mediante un sistema automatizado de pruebas de potencial de metano"	*Energies* (MDPI)	Investigar los posibles beneficios de aplicar una economía circular que convierta en energía la biomasa de los campos de refugiados sirios de Zaatari.	-	Empírica. El sistema automático de pruebas de potencial de metano (AMPTS) y los tubos graduados en la sala de aire acondicionado de temperatura controlada GB21. También se determinaron los valores caloríficos de los desechos orgánicos y los lodos, tanto en seco como en húmedo.	Siria. Laboratorio	La producción máxima de biogás a partir de un 100% de desechos orgánicos y un 100% de lodo utilizando AMPTS fue de 153 m3 ton-1 y 5,6 m3 ton-1, respectivamente. La producción de metano alcanzó su máximo en un rango de inóculo de 0,25 a 0,3 vs sub/vs. En cambio, la producción de metano disminuyó cuando el inóculo del vs sub/vs superó el 0,46.	Una posible aplicación de una economía circular en Zaatari y otros campos similares y pequeñas comunidades.
Asai, M.; Hayashi, T. y Yamamoto, M.	2019	"El análisis del modelo mental de las percepciones y políticas sobre la energía del biogás revela posibles limitaciones en una comunidad agrícola japonesa"	*Sustainability* (MDPI)	Analizar las percepciones de los interesados locales en el biogás utilizando un enfoque de modelo mental.	-	Empírica. Estudio de caso de veintidós personas. Modelo mental.	Comunidad en Japón	Se descubrió que muchos interesados compartían los beneficios cognitivos del biogás, mientras que había diferencias de percepción en cuanto al uso del digestato. Los productores agrícolas mencionaron restricciones técnicas y no técnicas a la aceptación del digestato, mientras que los productores lecheros y no lecheros se mostraron ambivalentes respecto de esas restricciones por el lado de la demanda.	Compartir los modelos mentales obtenidos entre los diversos interesados en un taller. Esto puede mejorar la comprensión y el aprendizaje social, y así apoyar mejor el establecimiento de sistemas de biogás sostenibles.

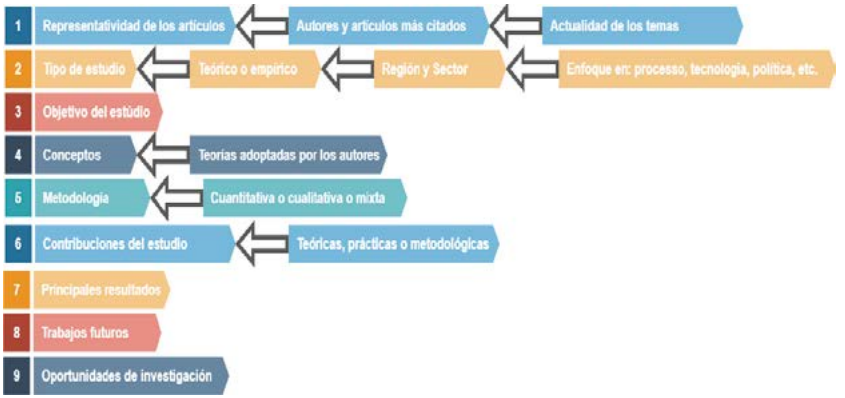
Fuente: elaboración propia

Luego el investigador puede realizar un análisis más detallado de estos artículos. Con la comprensión del área que se está investigando a través del estado del arte se puede realizar a continuación el análisis del contenido específico, creando preguntas más profundas para ser analizadas, buscando esas respuestas en los artículos seleccionados.

Por ejemplo, el investigador está llevando a cabo una investigación en el área de la simulación, por lo que necesita saber: 1) qué tipo de software se utiliza para esta simulación; 2) qué variables y/o indicadores se utilizan para esta simulación; 3) cuándo los parámetros utilizados en esta simulación son fórmulas matemáticas; 4) qué escenarios se realizan para este tipo de simulación, entre otros temas que el investigador puede decidir para buscar en los trabajos.

Estas oportunidades de investigación se construyen sobre los resultados encontrados en el modelo de la fase 4 del SYSMAP mediante un cuadro que formula preguntas y/o lentes teóricas sobre las principales categorías analizadas, como se muestra en el ejemplo de la figura 11 y el cuadro 6.

Figura 11. Fase 4: análisis de oportunidades de investigación/acceso



Fuente: elaboración propia (Rodrigues Vaz y Uriona-Maldonado, 2017)

Cuadro 6. Oportunidades de investigación

Categorías de evaluación/lentes	Preguntas de investigación
Objetivo de identificar la investigación que se desarrolló para el tema establecido por el investigador	¿Cómo introducimos la evaluación del rendimiento que estudiamos y practicamos sobre la gestión estratégica de la organización?
El tiempo y la geografía que desarrollará el investigador en relación con el tema	¿Cómo preparar un proceso de evaluación de la gestión estratégica de la organización que se renueve con el tiempo en un país determinado? ¿Y si este proceso es diferente de ciertos estados del mismo país?
La inclusión/exclusión de competencias y la toma de decisiones que debe desarrollar el investigador sobre el tema en cuestión	¿Cómo se desarrolla una evaluación del rendimiento que tenga en cuenta los valores y preferencias de la persona que toma las decisiones?
Metodología o formas de medición utilizadas para el sujeto establecidas por el investigador	¿Cómo medir el logro de los objetivos individuales y generales en un contexto específico de gestión estratégica de la organización?
Diagnóstico de la situación actual de la inserción e investigación a realizar por el investigador	¿Cómo diagnosticar la situación actual de la gestión estratégica organizacional en una determinada organización, tanto los instrumentos cualitativos para identificar y organizar los objetivos institucionales como los instrumentos cuantitativos para crear y priorizar acciones con un impacto más global y sistémico, según la percepción de sus directivos?
Mejora de las recomendaciones de los futuros trabajos de los autores investigados para el desarrollo del tema definido por el investigador	¿Cómo utilizar los conocimientos generados por estos instrumentos de evaluación del desarrollo para crear medidas que mejoren el logro de los objetivos de la gestión estratégica de la organización?

Fuente: elaboración propia (Rodrigues Vaz y Uriona-Maldonado, 2017)

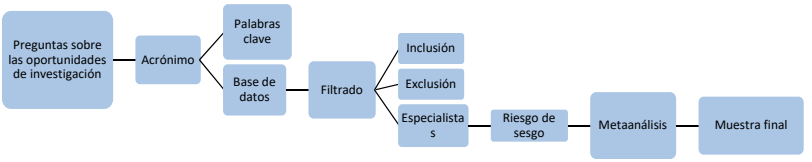
Metaanálisis

Después de la comprensión y el conocimiento del estado del arte sobre el tema investigado, el investigador puede continuar en la profundización del tema y realizar una segunda revisión tratándose ahora de una

revisión sistemática del contenido, con la repetición de algunos pasos y la inclusión de otros para su mejora.

De este modo, la investigación se inicia con algunas de las preguntas que se encuentran en las oportunidades de investigación, elegidas por el investigador, según su interés de investigación. Después se ajusta el acrónimo para la construcción adecuada de las palabras clave, para la búsqueda en las bases de datos de los artículos y con la ayuda de expertos en la materia realiza el filtrado de los artículos (decisión de inclusión y exclusión). A continuación, se evalúa el riesgo de sesgo de cada uno de los artículos seleccionados. Y, por último, se realiza el metaanálisis de los artículos que representarán la muestra final del verdadero resultado de las pruebas de conocimiento de esta brecha de investigación elegida. El diagrama 2 muestra el paso de reconstrucción de la revisión sistemática.

Diagrama 2. Segunda revisión sistemática



Fuente: elaboración propia

Un metaanálisis es un enfoque estadístico que combina resultados relevantes para responder una pregunta de investigación, utilizado dentro de una revisión sistemática. Este resultado se presenta siempre mediante el diagrama de bosque (*forest plot*).

La importancia del metaanálisis consiste principalmente en sintetizar las pruebas disponibles y señalar las áreas en las que es necesario seguir investigando. Así, el metaanálisis agrupa en un único resultado la media ponderada de cada variable descripta en los estudios primarios incluidos en la revisión sistemática (De Luca Canto, Massignan y Stefani, 2021).

Así pues, estos metaanálisis sirven para mostrar el verdadero resultado, principalmente entre los resultados divergentes (significativos frente a no significativos) de los estudios publicados sobre un tema determinado. Deberán reelaborarse cuando se publiquen más estudios que puedan cambiar el resultado real actual.

Los análisis estadísticos pueden abarcar desde las estadísticas básicas (media, desviación estándar, varianza) hasta la regresión, las pruebas de hipótesis, el análisis de la varianza (ANOVA), las ecuaciones estructurales y el análisis factorial, entre otros análisis cuantitativos.

El software para apoyar el método SYSMAP

Para realizar y ejecutar una revisión estructurada de la literatura, existen varios programas informáticos que ayudan a la gestión y visualización de datos, construyendo redes de colaboración entre autores, documentos y referencias.

Los programas de análisis científico que este trabajo presentará se dividen en dos categorías:

- Software de gestión de referencias: EndNote, Mendeley, Zotero, HistCite
- Software para la visualización y/o construcción de gráficos de análisis: Sci2, VOSviewer, SciMAT, NAILS, Gephi y CitNetExplorer

Cada uno de estos softwares tiene características especiales y, según el caso, estos programas pueden ser complementarios en la misma investigación, lo que hace que los datos sean más completos. A continuación, se presentarán estos diez softwares más usados. Vale la pena mencionar que existen otros softwares como CiteSpace, Vantage Point, BibExcel, INSPIRE que no serán revisados.

- HistCite es una implementación de software de historiografía algorítmica, desarrollado por Eugene Garfield, fundador del Instituto para la Información Científica e inventor del índice de citas. Es una herramienta flexible para ayudar a los investigadores a visualizar los resultados de la investigación bibliográfica, con una única interfaz vinculada con la base de datos de la Web of Science. Esta herramienta también permite analizar y organizar los resultados de una encuesta para obtener diversos puntos de vista sobre la estructura del tema, la historia y las relaciones (Garfield y Pudovkin, 2004).

- EndNote es un gestor de referencias bibliográficas producido por Thomson Scientific que trabaja integrado para la Web of Science. Facilita la labor de investigación, la redacción de documentos científicos y la recolección de referencias bibliográficas de las bases de datos en línea, importando los metadatos y agrupándolos de diferentes maneras (Gotschall, 2021).
- El software Mendeley permite generar estadísticas relacionadas con el número de artículos fundados, las regiones geográficas, la identificación de los lectores por zona, los autores que están investigando el tema de interés, entre otros. Mendeley trabaja con redes sociales, que permiten la interacción entre la comunidad y los responsables de la herramienta (Cauchik Miguel *et al.*, 2014).
- Zotero es un software desarrollado por la Universidad George Mason. Además de Mendeley y EndNote, Zotero puede investigar, almacenar y organizar las referencias bibliográficas obtenidas en bases de datos autorizadas (Echeverry-Mejía, Parano y Sánchez, 2022).
- VOSviewer es un software utilizado para construir redes bibliométricas basadas en datos descargados de bases de datos bibliográficas, como Web of Science y Scopus. El software ofrece al usuario la posibilidad de elegir entre los métodos de conteo total y fraccionario (Costas y Orduña-Malea, 2021; van Eck y Waltman, 2010).
- CitNetExplorer es una herramienta única de la Web of Science para construir redes de autores a través de la historiografía algorítmica. Se utiliza en la búsqueda de la historiografía de los autores y sus años de publicación (van Eck y Waltman, 2014).
- Network Analysis Interface for Literature Review (NAILS) es una herramienta de análisis de la literatura que utiliza una serie de funciones estadísticas y de análisis de redes habituales para proporcionar al usuario una visión general de los conjuntos de datos de la literatura de Web of Science (Hajikhani *et al.*, 2015).
- Science of Science (Sci2), a menudo citada en la literatura de los últimos dos años, es una herramienta diseñada para el estudio de la ciencia. Ofrece varias características para facilitar la comprensión e interpretación de las pautas: puntos de interés temporales, que determinan los períodos de máxima actividad en algún tema; identificación de nombres de lugares y asignación del código

geográfico correspondiente para asignarlos geográficamente; identificación de temas de interés, mediante grupos independientes, y también producción de redes de coautores (Li, Tian y Xu, 2021).

- SciMAT es un software de mapeo científico de descarga gratuita que incorpora métodos, algoritmos y mediciones generales para todas las etapas del flujo de trabajo del mapeo científico. También permite que el usuario realice varios estudios basados en redes bibliométricas y ofrece, en modo de visualización, tres representaciones simultáneas que permiten al usuario comprender mejor los resultados (Cobo *et al.*, 2012, 2021).
- Gephi es una plataforma interactiva para la visualización y exploración de todo tipo de redes y sistemas gráficos complejos, dinámicos y jerárquicos. Utiliza un mecanismo de renderización 3D para la visualización de redes en tiempo real, permite la visualización espacial y una fácil manipulación (Bastian, Heymann y Jacomy, 2009)

A modo de resumen, el cuadro 7 presenta las principales características y diferencias de cada uno de los programas informáticos mencionados anteriormente para el análisis bibliométrico.

Cuadro 7. Características de los softwares para el análisis bibliométrico/cienciométrico

EndNote	Mendeley	Zotero	HistCite	Herramientas de Sci2	VOSviewer	SciMAT	NAILS	CitNetExplorer	Gephi
Gestión y almacenamiento de referencias	Gestión y almacenamiento de referencias	Gestión y almacenamiento de referencias	Datos bibliométricos y visualización de historiogramas	Visualización de datos bibliométricos	Visualización de datos bibliométricos	Visualización de datos bibliométricos	Visualización de datos bibliométricos	Datos bibliométricos y visualización de historiogramas	Visualización de datos bibliométricos
En el escritorio y en línea	En el escritorio y en línea	En el escritorio y en línea	Escritorio	Escritorio	Escritorio	Escritorio	En línea	Escritorio	Escritorio
Mac, Windows	Mac, Windows	Mac, Windows	Windows	Mac, Windows	Mac, Windows	Mac, Windows	Mac, Windows	Mac, Windows	Mac, Windows
Se puede usar con todas las bases de datos y con Google Academic	Puede utilizarse con todas las bases de datos	Puede utilizarse con todas las bases de datos	Solo se puede usar con Web of Science (ISI)	Se puede usar con Scopus, ISI y RIS	Se puede usar con Scopus, ISI, formato RIS y csv	Se puede usar con Scopus, ISI, formato RIS y csv	Solo se puede usar con Web of Science (ISI)	Se puede usar con Scopus, ISI y RIS	Puede utilizarse con el formato SVG
Puede ser almacenado en .pdf	Puede ser almacenado en .pdf	Puede ser almacenado en .pdf	-	-	-	-	-	-	-
Puede ser compartido con otros (x8)	Puede ser compartido con otros	Puede ser compartido con otros	-	-	-	-	-	-	-
Se pueden crear otros campos de búsqueda	-	-	-	-	-	-	-	-	-
Se pueden ver los datos de los artículos (bibliografía temática) y exportarlos a un excel	-	-	Los datos pueden ser guardados en el software Gephi	Los datos pueden ser guardados en el software Gephi	-	-	Los datos pueden ser guardados en el software Gephi	-	-

Fuente: elaboración propia

Para el análisis de contenido este trabajo se sugiere el uso del software EndNote para gestionar y almacenar documentos PDF y la hoja de cálculo de Excel para organizar los datos recogidos después de su lectura y la construcción de las lagunas/oportunidades de investigación.

Existen también software como NVivo, ATLAS.ti, StArt, Vantage Point, MAXQDA, Orbit Intellixir, entre otros, que ayudan a organizar la información cualitativa del contenido de los artículos encontrados.

- NVivo: trabaja con el concepto del proyecto. Las fuentes de información de los proyectos, así como los datos generados durante el proceso de análisis y las categorías de información, se almacenan en una base de datos (Dhakal, 2022).
- ATLAS.ti: es una herramienta para el análisis de datos cualitativos que puede facilitar el manejo e interpretación de estos datos (Bach y Walter, 2015).
- StArt (State of the Art through Systematic Review): es una técnica utilizada para buscar evidencias en la literatura científica que se realiza de manera formal, aplicando pasos bien definidos, de acuerdo con un protocolo previamente elaborado, pasos y actividades; su ejecución es laboriosa y repetitiva (Fabbri *et al.*, 2010).
- Vantage Point: es una herramienta importante para la minería y el análisis de datos. A través de este software es posible realizar análisis y estudios relacionados con la producción científica, la producción docente, la producción de tesis y disertaciones, el análisis de colecciones, el análisis temático, así como el análisis de las tendencias de investigación (Dudziak, 2014).
- Orbit Intellixir: es un sistema de búsqueda, selección, análisis y exportación de información contenida en las patentes (Flynn *et al.*, 2020).
- MAXQDA: es un software académico para el análisis de datos cualitativos y métodos de investigación mixtos, como el análisis de contenido, entrevistas, discursos, grupos de discusión, archivos de audio/video/imágenes, datos de Twitter, entre muchas otras posibilidades (Saillard, 2011).

Debate sobre la información y los datos necesarios para la aplicación del método en el CITS

En general, los estudios e investigaciones realizados en América Latina no están en las fronteras del conocimiento científico mundial, debido a la falta de herramientas y recursos que no llegan a los estudios e investigaciones existentes y disponibles en los países desarrollados. Así pues, la utilización de métodos estructurados de examen de la bibliografía puede servir para identificar las lagunas de la investigación o las lagunas más pertinentes para el investigador, a fin de reducir la distancia científica entre la investigación de América Latina y la de otras regiones del mundo.

Sin embargo, hay por lo menos tres factores que deben ser discutidos. Primero: el acceso a los datos necesarios para aplicar el método SYSMAP. Para ello, es importante recordar brevemente el modelo de gestión editorial científica dominante hoy en día. Este modelo es predominantemente pagado, lo que significa que la mayoría de las revistas científicas más reconocidas del mundo ofrecen acceso restricto a los suscriptores (por lo general universidades) a través de editoriales científicas, entre las que se puede citar, por ejemplo, la editorial académica Elsevier. A esto se añade el modelo de base de datos en línea, que reúne agrupaciones de revistas científicas de diferentes editoriales.

Volviendo a nuestro primer factor, el acceso a los datos requiere que la institución a la que pertenece el investigador tenga acceso a las bases de datos, por ejemplo, Web of Science de Thompson Reuters y Scopus de Elsevier.

En América Latina, solo unos pocos países tienen políticas de CTI que apoyan el acceso a las bases de datos internacionales. En el caso de Brasil, por ejemplo, el organismo CAPES gastó aproximadamente cien millones de dólares en 2016 en la suscripción a bases de datos (CAPES, 2016) para que todas las instituciones educativas federales pudieran acceder a las fuentes científicas. Considerando que existen aproximadamente 108 instituciones educativas federales (Inep, 2018), el monto gastado por cada institución fue cercano a los novecientos mil dólares. Sin embargo, si se compara con las instituciones de los países desarrollados, este gasto sigue siendo pequeño. La Universidad de Harvard, por ejemplo, gastó tres mil doscientos millones de dólares para acceder a las bases de datos internacionales (Sample, 2012), casi cuatro veces más en términos relativos.

Una suscripción más cara también significa un acceso más amplio,[2] en términos de período de tiempo (acceso a toda la colección de una determinada revista científica desde su creación), temático (más revistas en cada área temática) y de inmediatez (*early access* para los artículos recientemente publicados, cuando lo contrario significa esperar un período de embargo que hace imposible el acceso a los artículos de los últimos uno o dos años, por ejemplo).

Un segundo aspecto que debe examinarse, estrechamente relacionado con el anterior, es la disponibilidad de software para los estudios de revisión de la literatura estructurada. Debido a las características del actual modelo comercial de acceso a bases de datos científicas, los desarrolladores de software siguen una externalidad de red, es decir, producen software que funciona con las bases que más cuentan, a fin de disminuir sus propios riesgos.

La operación se debe principalmente a la posibilidad de exportar metadatos (archivos que contienen información bibliográfica) de la base de datos[3] al programa informático de análisis. Este es el caso de VOSviewer, HistCite, Orbit Intellixir y Vantage Point, entre otros.

Esto significa que si el investigador no tiene acceso a la base de datos, también perderá la posibilidad de utilizar programas informáticos modernos para el contenido y el análisis científico. Por otra parte, las funcionalidades también son superiores en los programas informáticos pagos , como Orbit Intellixir y Vantage Point, que permiten realizar una serie de análisis complejos, no disponible en los programas informáticos gratuitos competidores o muy difíciles de reproducir.

Finalmente, un tercer aspecto es el uso del inglés como la *lingua franca* de la ciencia. Prácticamente toda la literatura científica de vanguardia está escrita en inglés, lo que dificulta el acceso a esos materiales si el investigador no domina el idioma, así como la publicación de trabajos científicos latinoamericanos (a menudo impedidos en la etapa de revisión por pares por deficiencias en la escritura en inglés). Incluso, muchos de los programas de análisis están disponibles solo en inglés.

A pesar de los aspectos mencionados, se observa con el uso de métodos estructurados de revisión de la literatura la posibilidad de superar estas limitaciones, al menos parcialmente. El uso de bases de datos de

2 Este aspecto lo experimentó uno de los autores del capítulo cuando estudiaba en la Universidad de Duke.
3 Prácticamente todo el software funciona solo con Web of Science.

libre acceso se ha fomentado en América Latina durante por lo menos un decenio.

La Scientific Electronic Library Online (SciELO) nació de los esfuerzos realizados por la Fundación de Apoyo a la Investigación del Estado de São Paulo (FAPESP) y el Centro Latinoamericano y del Caribe de Información en Ciencias de la Salud (BIREME) (Santos, 2003) con el objetivo de fomentar la producción científica latinoamericana, aumentando su visibilidad (siguiendo los más altos estándares de gestión de la información, incluyendo la facilitación de la generación y exportación de metadatos para el análisis bibliométrico).

En la actualidad, SciELO ya cuenta con portales para casi todos los países de la región, agrupando lo mejor de la producción científica de cada país, gracias a los altos estándares requeridos para las revistas que desean ser indexadas en la base. El uso de SciELO para estudios de revisión estructurada de la literatura es posible gracias al mecanismo de exportación de metadatos, en formatos típicos de otras bases internacionales.

Herramientas como EndNote, Zotero y otras pueden leer automáticamente estos archivos exportados para facilitar el análisis de la información extraída.

En la misma línea, también hay iniciativas que fomentan el modelo de revista de acceso abierto. El más conocido es el modelo del Public Knowledge Project (PKP) (Simon Fraser Universtiy, 2020) que fomenta el uso de plataformas de gestión editorial de acceso abierto. De esta manera, las revistas que no dependen de los recursos de las principales editoriales científicas pueden generar sus procesos editoriales con las herramientas de PKP, especialmente el Open Journal Systems (OJS),, para las revistas y el Open Conference Systems (OCS), para la gestión de los congresos científicos.

En Brasil, por ejemplo, ha habido un creciente incentivo para crear revistas de acceso abierto apoyado por líneas de financiación pública para ayudar a mantener las actividades editoriales (Medina Kern y Urriona-Maldonado, 2018). Esto, junto con la inclusión de estas revistas en los procesos de evaluación de investigadores y programas de posgrado, en relación con la medición de las publicaciones, condujo a un aumento de las revistas brasileñas de acceso abierto.

Uno de los resultados de este proceso ha sido el aumento de las revistas de libre acceso en las bases internacionales por suscripción. Un estudio realizado por Borges de Oliveira y Schwarz Rodrigues (2012) identificó que de las 556 revistas latinoamericanas indexadas en Web of

Science y Scopus, el 98% era de acceso abierto y el 69% también estaba indexado en SciELO.

Esto lleva a la conclusión de que muchas revistas latinoamericanas se sometieron primero a la indización en la base de datos SciELO, lo que generó aptitudes editoriales y publicaciones de mayor calidad, ayudando más tarde a la indización en bases de datos de pago, como Web of Science y Scopus.

Por último, también hay oportunidades en relación con el uso de software libre para el análisis de la literatura. Las aplicaciones y paquetes de software R y Python pueden realizar análisis de redes y minería de textos similares a VOSviewer, Gephi y NVivo, con mayor agilidad y sin las restricciones de los metadatos procedentes de la Web of Science o Scopus.

El crecimiento de las aplicaciones R y Python es orgánico y está impulsado por comunidades globales de usuarios y desarrolladores. El paquete Bibliometrix (Aria y Cuccurullo, 2017) es un ejemplo de aplicaciones centradas en el análisis bibliométrico, que incluye el análisis de citas, cocitación, redes e historiogramas y la integración de bases de datos (Caputo y Kargina, 2022).

Recomendaciones para futuras aplicaciones

En primer lugar, el método SYSMAP puede aplicarse en cualquier área de conocimiento, con cualquier base de datos, y puede utilizarse con la combinación de diversos programas informáticos.

Las recomendaciones de aplicación comienzan con la elección adecuada de las palabras clave, porque una vez que se eligen las equivocadas, toda la investigación se ve comprometida y los datos no aportarán resultados y pruebas significativas al tema investigado. Por lo tanto, se deben identificar todos los sinónimos y acrónimos posibles antes de iniciar la búsqueda. También se debe hacer el proceso de búsqueda varias veces para estar seguro del resultado encontrado.

Otra recomendación de aplicación es en relación con las bases de datos, en las que es importante que el investigador tenga un conocimiento adecuado del campo de estudio, es decir, que tenga una noción de las bases de datos más importantes en su área, como por ejemplo, en ingeniería la Web of Science, Scopus, ScienceDirect; en Salud, PubMed, LILACS, BDENF; en ciencias sociales, la SSCI de la Web of Science, entre otras. El acceso limitado de la universidad que trabaja con las bases de

datos es un tema delicado, pero el investigador puede tener otro acceso al artículo, como el correo electrónico directo del autor, ResearchGate, Twitter, Facebook, entre otras plataformas digitales.

En cuanto a los elementos de inclusión y exclusión de artículos en la muestra (fase de filtrado), si el investigador no tiene algún cuidado, puede comprometer toda la investigación, porque si se realiza un filtrado muy restringido puede dejar de lado artículos importantes y, si se deja muy abierto, puede generar un volumen de artículos innecesarios y difíciles de asimilar. Así pues, es necesario crear elementos específicos para la exclusión de los artículos y que se especifiquen los criterios. La inclusión debe ser llevada a cabo de la misma manera que la exclusión.

Para el análisis científico es importante que se realicen los análisis de todos los artículos, incluidos aquellos que forman parte de la "fase de filtrado", para que el investigador pueda disponer de la población de la muestra. Otra recomendación pertinente de esta fase es que el investigador no realice una sola forma de análisis de citas o cotizaciones, sino que combine varias citas, cocitaciones, referencias, hitoriogramas, espaciales, con la ayuda de diversos tipos de programas informáticos, para una mayor comprensión del área temática investigada.

Y por último, realizar el análisis de contenido, esta recomendación es necesaria para buscar artículos en todas las plataformas digitales (Google, Google Scholar, Twitter, ResearchGate, correo electrónico personal, Facebook, entre otros). El investigador debe realizar una lectura en profundidad del artículo siguiendo los criterios de análisis. Se recomienda que más de una persona realice este análisis de contenido para que se consideren todos los aspectos en el análisis.

Bibliografía

Aisenberg Ferenhof, H. y Fernandes, R. F. (2016). "Desmistificando a revisão de literatura como base para redação científica: método SFF". *Revista ACB*, vol. 21, nº 3, pp. 550-563.

Akl, E. A.; Bossuyt, P. M.; Boutron, I.; Brennan, S. E.; Chou, R.; Glanville, J.; Grimshaw, J. M.; Hoffmann, T. C.; Hróbjartsson, A.; Lalu, M. M.; Li, T.; Loder, E. W.; Mayo-Wilson, E.; McDonald, S.; McGuinness, L. A.; McKenzie, J. E.; Moher, D.; Mulrow, C. D.; Page, M. J.; Shamseer, L.; Stewart, L. A.; Tetzlaff, J. M.; Thomas, J.; Tricco, A. C.;

Welch, V. A. y Whiting, P. (2021). "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews". *BMJ*, vol. 372, nº 71. DOI: 10.1136/bmj.n71.

Alves Anacleto, C.; de Souza Mendonça, A. K.; Pacheco Paladini, E.; Rodrigues Vaz, C. y Rojas Lezana, Á. G. (2017). "Comparing Patent and Scientific Literature in Airborne Wind Energy". *Sustainability*, vol. 9, nº 6.

Antes, G.; Khan, K. S.; Kleijnen, J. y Kunz, R. (2003). "Five steps to conducting a systematic review". *JRSM: Journal of the royal society of medicine*, vol. 96, nº 3, pp. 118-121.

Aria, M. y Cuccurullo, C. (2017). "bibliometrix: An R-tool for comprehensive science mapping analysis". *Journal of Informetrics*, vol. 11, nº 4, pp. 959-975. DOI: 10.1016/j.joi.2017.08.007.

Bach, T. M. y Walter, S. A. (2015). "Adeus papel, marca-textos, tesoura ecola: inovando o processo de análise de conteúdo por meio do Atlas. ti". *Administração: ensino e pesquisa*, vol. 16, nº 2, pp. 275-308.

Bardin, L. (2011). *Análise de conteúdo*. São Paulo: Edições 70.

Barratt, M.; Choi, T. Y. y Li, M. (2011). "Qualitative case studies in operations management: Trends, research outcomes, and future research implications". *Journal of Operations Management*, vol. 29, nº 4, pp. 329-342.

Bastian, M.; Heymann, S. y Jacomy, M. (2009). "Gephi: an open source software for exploring and manipulating networks". *ICWSM*, vol. 3, nº 1, 361-362.

Benneworth, P.; Coenen, L. y Truffer, B. (2012). "Toward a spatial perspective on sustainability transitions". *Research Policy*, vol. 41, nº 6, pp. 968-979. DOI: 10.1016/j.respol.2012.02.014.

Bergek, A.; Carlsson, B.; Jacobsson, S.; Lindmark, S. y Rickne, A. (2008). "Analyzing the functional dynamics of technological innovation systems: A scheme of analysis". *Research Policy*, vol. 37, nº 3, 407-429. DOI: 10.1016/j.respol.2007.12.003.

Binz, C.; Coenen, L. y Truffer, B. (2014). "Why space matters in technological innovation systems-Mapping global knowledge dynamics of membrane bioreactor technology". *Research Policy*, vol. 43, nº 1, 138-155. DOI: 10.1016/j.respol.2013.07.002.

Borges de Oliveira, A. y Schwarz Rodrigues, R. (2012). "Periódicos científicos na America Latina: títulos em Acesso Aberto indexados no ISI e SCOPUS". *Perspectivas em Ciência da Informação*, vol. 17, nº 4, pp. 77-99. DOI: https://doi.org/10.1590/S1413-99362012000400006.

Bornia, A. C.; Tezza, R. y Vey, I. H. (2010). "Sistemas de medição de desempenho: uma revisão e classificação da literatura". *Gestão & Produção*, vol. 17, nº 1, pp. 75-93.

Burgess, K.; Koroglu, R. y Singh, P. J (2006). "Supply chain management: a structured literature review and implications for future research". *International Journal of Operations & Production Management*, vol. 26, nº 7, pp. 703-729. DOI: https://doi.org/10.1108/01443570610672202.

Caputo, A. y Kargina, M. (2022). "A user-friendly method to merge Scopus and Web of Science data during bibliometric analysis". *Journal of Marketing Analytics*, vol. 10, nº 1, pp. 82-88. DOI: 10.1057/s41270-021-00142-7.

Carlsson, B. y Stankiewicz, R. (1991). "On the nature, function and composition of technological systems". *Journal of Evolutionary Economics*, vol. 1, nº 2, pp. 93-118.

Cauchik Miguel, P. A.; Hansch Beuren, F.; Issao Kubota, F.; Kazumi Yamakawa, E. y Scalvenzi, L. (2014). "Comparativo dos softwares de gerenciamento de referências bibliográficas: Mendeley, EndNote e Zotero". *Transinformação*, vol. 26, nº 2.

Chittó Stumpf, I. R. y de Souza Vanz, S. A. (2010). "Procedimentos e ferramentas aplicados aos estudos bibliométricos". *Informação & Sociedade*, vol. 20, nº 2.

Cobo, M. J.; Gamboa-Rosales, N. K.; Gutiérrez-Salcedo, M.; Herrera-Viedma, E.; López-Robles, J. R. y Martínez-Sánchez, M. A. (2021). "30th Anniversary of Applied Intelligence: A combination of bibliometrics and thematic analysis using SciMAT". *Applied Intelligence*, vol. 51, nº 9, pp. 6547-6568. DOI: 10.1007/s10489-021-02584-z.

Cobo, M. J.; Herrera, F.; Herrera-Viedma, E. y López-Herrera, A. G. (2012). "SciMAT: A new science mapping analysis software tool". *Journal of the American Society for Information Science and Technology*, vol. 63, nº 8, pp. 1609-1630. DOI: 10.1002/asi.22688.

Coenen, L. y Díaz López, F. J. (2010). "Comparing systems approaches to innovation and technological change for sustainable and competitive economies: an explorative study into conceptual commonalities, differences and complementarities". *Journal of Cleaner Production*, vol. 18, nº 12, pp. 1149-1160. DOI: 10.1016/j.jclepro.2010.04.003.

Colford Jr., J. M.; Enanoria, W.; Gorman, J. D.; Kennedy, G.; McCulloch, M.; Pai, M.; Pai, N. y Tharyan, P. (2003). "Systematic reviews and meta-analyses: an illustrated, step-by-step guide". *The National medical journal of India*, vol. 17, nº 2, pp. 86-95.

Colicchia, C. y Strozzi, F. (2012). "Supply chain risk management: a new methodology for a systematic literature review". *Supply Chain Management*, vol. 17, nº 4, pp. 403-418. DOI: https://doi.org/10.1108/13598541211246558.

Cooper, H.; Hedges, L. V. y Valentine, J. C. (eds.) (2009). *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation.

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) (2016). "Portal de Periódicos da CAPES: um patrimônio a ser preservado". *Gov.br - Ministério da Educação*. Disponible em: https://www.periodicos.capes.gov.br/?option=com_pnews&component=Clipping&view=pnewsclipping&cid=557&mn=0.

Costas, R. y Orduña-Malea, E. (2021). "Link-based approach to study scientific software usage: the case of VOSviewer". *Scientometrics*, vol. 126, nº 9, pp. 8153-8186. DOI: 10.1007/s11192-021-04082-y.

da Rosa, I. O.; Ensslin, L. y Rolim Ensslin, S. (2011). "Processo para seleção do referencial teórico para o gerenciamento de riscos afins à segurança empresarial". *Pesquisa & Desenvolvimento Engenharia de Produção*, vol. 9, nº 2, pp. 67-80.

da Silva, M. R.; Massao Hayashi, C. R. y Piumbato Innocentini Hayashi, M. C. (2011). "Análise bibliométrica e cientométrica: desafios para especialistas que atuam no campo". *INCID: Revista de Ciência da Informação e Documentação*, vol. 2, nº 1, pp. 110-129.

De Luca Canto, G. (2020). *Revisões sistemáticas da literatura: guia prático*. Curitiba: Brazil Publishing.

De Luca Canto, G.; Massignan, C. y Stefani, C. M. (2021). *Risco de viés em revisões sistemáticas: guia prático*. Curitiba: Brazil Publishing.

de Mesquita, M. A. y do Rego, J. R. (2011). "Controle de estoque de peças de reposição em local único: uma revisão da literatura". *Produção*, vol. 21, nº 4, pp. 645-655. DOI: 10.1590/S0103-65132011005000002

de Oliveira Lacerda, R. T.; Ensslin, L. y Rolim Ensslin, S. (2012). "Uma análise bibliométrica da literatura sobre estratégia e avaliação de desempenho". *Gestão & Produção*, vol. 19, nº 1.

Denyer, D.; Smart, P. y Tranfield, D. (2003). "Towards a methodology for developing evidence-informed management knowledge by means of systematic review". *British journal of management*, vol. 14, nº 3, pp. 207-222.

Dhakal, K. (2022). "NVivo". *JMLA: Journal of the Medical Library Association*, vol. 110, nº 2, pp. 270-272. DOI: 10.5195/jmla.2022.1271.

Donthu, N.; Kumar, S.; Lim, W. M.; Mukherjee, D. y Pandey, N. (2021). "How to conduct a bibliometric analysis: An overview and guidelines". *Journal of Business Research*, vol. 133, pp. 285-296. DOI: https://doi.org/10.1016/j.jbusres.2021.04.070.

Dosi, G. (1982). "Technological Paradigms and Technological Trajectories - A Suggested Interpretation of the determinants and directions of Technical Change". *Research Policy*, vol. 11, nº 3, pp. 147-162.

Dresch, A.; Pacheco Lacerda, D. y Valle Júnior Antunes, J. A. (2015). *Design science research: método de pesquisa para avanço da ciência e tecnologia*. Madrid: Bookman.

Dudziak, E. A. (coord.) (2014). *Manual de uso do VantagePoint*. São Paulo: Grupo de Estudos Bibliométricos Aplicados do SIBiUSP.

Echeverry-Mejía, J. A.; Parano, M. y Sánchez, H. J. (2022). "Zotero. Tu asistente personal de investigación y estudio. Guía para estudiantes y docentes". *Cuadernos de Coyuntura*, vol. 7, pp. 1-14.

Egger, M.; Higgins, J. P. y Smith, G. D. (2022). *Systematic Reviews in Health Research: Meta-Analysis in Context*. New York: John Wiley & Sons.

Ensslin, L.; Hartmann Feyh, M.; Rolim Ensslin, S. y Vieira de Souza, A. J. (2012). "Como construir conhecimento sobre o tema de pesquisa? Aplicação do processo ProKnow-C na busca de literatura

sobre avaliação do desenvolvimento sustentável". *Revista de Gestão Social e Ambiental*, vol. 5, nº 2, pp. 47-62.

Fabbri, S.; Hernandes, E. C. M.; Thommazo, A. y Zamboni, A. (2010). "StArt uma ferramenta computacional de apoio à revisão sistemática". Presentado en el *Congresso Brasileiro de Software (CBSoft'10)*. Salvador, Brasil.

Faria Fernandes, F. C. y Godinho Filho, M. (2007). "Sistemas de coordenação de ordens: revisão, classificação, funcionamento e aplicabilidade". *Revista Gestão & Produção, São Carlos*, vol. 14, nº 2.

Flynn, K. J.; Fuentes-Grünewald, C.; Gonçalves de Oliveira Jr., R.; Nicolau, E.; Picot, L. y Rumin, J. (2020). "A Bibliometric Analysis of Microalgae Research in the World, Europe, and the European Atlantic Area". *Marine Drugs*, vol. 8, nº 2, 79.

Fuenfschilling, L. y Truffer, B. (2014). "The structuration of socio-technical regimes - Conceptual foundations from institutional theory". *Research Policy*, vol. 43, nº 4, pp. 772-791. DOI: 10.1016/j.respol.2013.10.010.

Garfield, E. (2006). "Citation indexes for science. A new dimension in documentation through association of ideas". *International journal of epidemiology*, vol. 35, nº 5, pp. 1123-1127.

_____ (2009). "From the science of science to Scientometrics visualizing the history of science with HistCite software". *Journal of Informetrics*, vol. 3, nº 3, pp. 173-179. DOI: http://dx.doi.org/10.1016/j.joi.2009.03.009.

Garfield, E. y Pudovkin, A. I. (2004). "The HistCite system for mapping and bibliometric analysis of the output of searches using the ISI Web of Knowledge". Presentado en el *Proceedings of the 67th Annual Meeting of the American Society for Information Science and Technology*. Newport, Rhode Island, Estados Unidos.

Godinho Filho, M. y Lage Junior, M. (2008). "Adaptações ao sistema kanban: revisão, classificação, análise e avaliação". *Gestão & Produção*, vol. 15, nº 1, pp. 173-188.

Gohr, C. F.; Goncalves, A. M. C.; Pinto, N. O. y Santos, L. C. (2013). "Um método para a revisão sistemática da literatura em pesquisas de engenharia de produção". Presentado en el *XXXIII Encontro Nacional de Engenharia de Produçao*. Salvador, BA, Brasil.

Gotschall, T. (2021). "EndNote 20 desktop version". *JML: Journal of the Medical Library Asociation*, vol. 109, nº 3, pp. 520-522. DOI: 10.5195/jmla.2021.1260.

Gough, D.; Oliver, S. y Thomas, J. (2019). "Clarifying differences between reviews within evidence ecosystems". *Systematic reviews*, vol. 8, nº 1, pp. 1-15.

Hajikhani, A.; Ikonen, J.; Knutas, A.; Porras, J. y Salminen, J. (2015). "Cloud-based bibliometric analysis service for systematic mapping studies". Presentado en el *Proceedings of the 16th International Conference on Computer Systems and Technologies*. Dublín, Irlanda.

Hekkert, M. P.; Kuhlmann, S.; Negro, S. O.; Smits, R. E. H. M. y Suurs, R. A. A. (2007). "Functions of innovation systems: A new approach for analysing technological change". *Technological forecasting and social change*, vol. 74, nº 4, pp. 413-432. DOI: http://dx.doi.org/10.1016/j.techfore.2006.03.002.

Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep). (2018). *Censo da Educação Superior*. Brasilia: Inep. Disponible en: https://download.inep.gov.br/educacao_superior/censo_superior/documentos/2019/censo_da_educacao_superior_2018-notas_estatisticas.pdf.

Jacobsson, S. y Johnson, A. (2000). "The diffusion of renewable energy technology: an analytical framework and key issues for research". *Energy Policy*, vol. 28, nº 9, pp. 625-640. DOI: http://dx.doi.org/10.1016/S0301-4215(00)00041-0

Kovaleski, J. L.; Negri Pagani, R. y Resende, L. M. (2015). "Methodi Ordinatio: a proposed methodology to select and rank relevant scientific papers encompassing the impact factor, number of citation, and year of publication". *Scientometrics*, vol. 105, nº 3, pp. 2109-2135.

Lawhon, M., & Murphy, J. T. (2012). Socio-technical regimes and sustainability transitions: Insights from political ecology. *Progress in human geography*, *36*(3), 354-378.

Li, P.; Tian, S. y Xu, X. (2021). "Acknowledgement network and citation count: the moderating role of collaboration network". *Scientometrics*, vol. 126, nº 9, pp. 7837-7857. DOI: 10.1007/s11192-021-04090-y.

Lundvall, B.-Å. (ed.) (1992). *National Systems of Innovation: towards a theory of innovation and interactive learning*. Londres: Pinter.

Macedo dos Santos, R. N.; Silva Santos, J. L. yUriona-Maldonado, M. (2010). "Inovação e conhecimento organizacional: um mapeamento bibliométrico das publicações científicas até 2009". Presentado en el *XXXIV encontro da ANPAD*. Rio de Janeiro, Brasil.

Markard, J. y Truffer, B. (2008). "Technological innovation systems and the multi-level perspective: Towards an integrated framework". *Research Policy*, vol. 37, nº 4, pp. 596-615. DOI: 10.1016/j.respol.2008.01.004.

Markard, J.; Raven, R. y Truffer, B. (2012). "Sustainability transitions: An emerging field of research and its prospects". *Research Policy*, vol. 41, nº 6, pp. 955-967. DOI: 10.1016/j.respol.2012.02.013.

Medina Kern, V. y Uriona-Maldonado, M. (2018). "Cenários da dinâmica de hipercrescimento e colapso das revistas científicas brasileiras líderes na Ciência da Informação". *Em Questão*, vol. 24, pp. 258-277.

Mulchenko, Z. M. y Nalimov, V. V. (1971). *Measurement of Science. Study of the Development of Science as an Information Process*. Springfield, Virginia: National Technical Information Service.

Musiolik, J., Markard, J., & Hekkert, M. (2012). Networks and network resources in technological innovation systems: Towards a conceptual framework for system building. *Technological Forecasting and Social Change*, *79*(6), 1032-1048.

Nelson, R. R. y Winter, S. (1982). *An evolutionary theory of economic change*. Cambridge: Harvard University Press.

Oliveira Araújo, W. C. (2020). "Recuperação da informação em saúde: construção, modelos e estratégias". *ConCI: Convergências em Ciência da Informação*, vol. 3, nº 2, pp. 100-134. DOI: 10.33467/conci.v3i2.13447.

Pacheco Lacerda, D. y Teixeira, R. (2010). "Gestão da cadeia de suprimentos: análise dos artigos publicados em alguns periódicos acadêmicos entre os anos de 2004 e 2006". *Gestão & Produção*, vol. 17, nº 1, pp. 207-227.

Pritchard, A. (1969). "Statistical bibliography or bibliometrics?". *Journal of Documentation*, vol. 25, n° 4, pp. 348-349.

Raven, R. y Smith, A. (2012). "What is protective space? Reconsidering niches in transitions to sustainability". *Research Policy*, vol. 41, n° 6, pp. 1025-1036. DOI: 10.1016/j.respol.2011.12.012.

Rodrigues Vaz, C. y Uriona-Maldonado, M. (2015). "The evolution of sustainability transitions and technological innovation systems research: a bibliometric analysis". Presentado en el *15th Globelics International Conference*. Atenas, Grecia.

_____ (eds.) (2017). "Revisão de literatura estruturada: proposta do modelo SYSMAP (Scientometric and Systematic Yielding Mapping Process)". *Aplicações de Bibliometria e Análise de Conteúdo em casos da Engenharia de Produção*, pp. 21-42. Brasil: Universidade Federal de Santa Catarina (UFSC).

Røstvik, H. N. (2018). "The mobility revolution as seen through Norwegian eyes". *Architectural Science Review*, vol. 61, n° 5, pp. 362-366. DOI: 10.1080/00038628.2018.1502152.

Saillard, E. K. (2011). "Systematic versus interpretive analysis with two CAQDAS packages: NVivo and MAXQDA". *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, vol. 12, n° 1.

Sample, I. (24 de abril de 2012). "Harvard University says it can't afford journal publishers' prices". *The Guardian*. Disponible en: https://www.theguardian.com/science/2012/apr/24/harvard-university-journal-publishers-prices.

Simon Fraser Universtiy (2020). *Public Knowledge Project*. URL: https://pkp.sfu.ca/.

Smith, A., Voß, J. P., & Grin, J. (2010). Innovation studies and sustainability transitions: The allure of the multi-level perspective and its challenges. *Research policy*, *39*(4), 435-448.

Suurs, R. A., & Hekkert, M. P. (2009). Cumulative causation in the formation of a technological innovation system: The case of biofuels in the Netherlands. *Technological Forecasting and Social Change*, *76*(8), 1003-1020.

van Eck, N. J. y Waltman, L. (2010). "Software survey: VOSviewer, a computer program for bibliometric mapping". *Scientometrics*, vol. 84, nº 2, pp. 523-538. DOI: 10.1007/s11192-009-0146-3.

_____ (2014). "CitNetExplorer: A new software tool for analyzing and visualizing citation networks". *Journal of Informetrics*, vol. 8, nº 4, pp. 802-823.

Capítulo 6
Modelación y simulación como herramientas para la comprensión de fenómenos emergentes en la difusión y transferencia de tecnologías

William Alejandro Orjuela Garzón, Santiago Quintero Ramírez

Introducción

La transferencia y difusión de tecnología es un fenómeno social complejo en el cual las capacidades, la interacción y toma de decisiones de los agentes vinculados a los procesos de generación, difusión y uso de la tecnología (Carlsson y Stankiewicz, 1991) son un factor crítico para garantizar su éxito. Dadas las características de dicho fenómeno, los métodos matemáticos empleados para su estudio y análisis son estáticos y no introducen características dinámicas y heterogéneas, que varían en el tiempo (Günther *et al.*, 2012). A partir de estas características, se ha identificado el potencial de los paradigmas de modelación y simulación para mejorar su comprensión y desempeño, esto debido a su utilidad para simular los problemas del mundo real con mayor precisión (Gilbert, 2007) y tener en cuenta las características de los agentes, sus comportamientos, la heterogeneidad y la emergencia de patrones en el macronivel.

El modelado y la simulación tienen el objetivo de reproducir e imitar el comportamiento de un sistema en la vida real (Guerrero, Schwarz y Slinger, 2016). La modelación basada en agentes (en adelante, MBA) es una técnica formada por las ciencias sociales computacionales que involucra la construcción de modelos que son programas de computador (Gilbert,

2007) compuestos por agentes autónomos que tienen comportamientos descriptos por reglas simples, a su vez, estos interactúan con otros agentes, los cuales influencian sus comportamientos (Macal y North, 2010). La MBA pertenece al enfoque de modelado de "abajo hacia arriba", no hay un planificador central que controle el sistema como un todo y para ese efecto controla el comportamiento de los agentes individuales en el nivel agregado (Happe, 2004).

Es decir, las propiedades del macronivel solo pueden ser entendidas propiamente como un resultado de las microdinámicas evolutivas de los agentes (Tesfatsion, 2002), por lo tanto, las propiedades agregadas que emergen son interpretadas como el resultado de interacciones repetidas entre agentes, estos patrones, comportamientos y estructuras que emergen no son explícitamente programados en el modelo, sino que surgen de dichas interacciones. Estos agentes interactúan en el medio, a través de toma de decisiones que dependen de las expectativas de beneficio, la influencia de elecciones pasadas echas por otros agentes y las dinámicas de decisión preestablecidas para cada tipo de agente en el modelo (Fagiolo, Moneta y Windrum, 2007).

Esta técnica ha sido ampliamente usada en las ciencias sociales ya que involucra la creación de modelos que contemplan agentes heterogéneos, cada uno de los cuales representa un actor social y un ambiente en el cual los agentes interactúan y se influencian mutuamente, aprenden de sus experiencias y adaptan sus comportamientos a su entorno (Macal y North, 2010). Estos agentes que interactúan entre ellos son programados para ser proactivos, autónomos y capaces de percibir el mundo virtual (Gilbert, 2008). La MBA ha sido empleada en el estudio de fenómenos sociales complejos como dinámicas de opinión, adopción de tecnologías, mercadeo, redes de innovación, estudios organizacionales, diseño de política (Elsenbroich y Gilbert, 2014), aprendizaje y desaprendizaje en sistemas de innovación (Quintero, 2016), puesto que permite establecer escenarios que facilitan la comprensión de los fenómenos a través del tiempo, dejando a un lado la mirada estática de otros marcos analíticos.

La modelación basada en agentes ha venido ganando interés en el ámbito científico en los últimos veinte años, y muestra de ello es la creciente evolución en la producción de artículos de investigación en revistas académicas, que a la fecha es cercana a los diecisiete mil según datos de Scopus.[1]

1 Base de datos consultada en enero de 2020, bajo el siguiente algoritmo: (TITLE-ABS-KEY (("agent-based model*") OR ("ABM"))).

Pasando de tan solo sesenta y ocho documentos publicados en el año 2000 a cerca de mil seiscientos en el 2019. Una de las principales revistas académicas es *Lecture Notes in Computer Science* con cerca de 655 publicaciones en el área de la MBA, seguido por *The Journal of Artificial Societies and Social Simulation* (*JASS*, por sus siglas en inglés), de la universidad de Surrey con cerca de 367 publicaciones relacionadas con la MBA. Entre los autores con mayor número de publicaciones se encuentra Uri Wilensky con cuarenta y tres documentos, resaltando que Wilensky fue el desarrollador de uno de los software más empleados en la MBA (NetLogo) (Wilensky, 1999).

El objetivo de este capítulo es mostrar las diferentes aplicaciones que puede tener la MBA, para mejorar la compresión de los fenómenos de CTI en Latinoamérica, haciendo un énfasis especial en la transferencia y difusión de tecnologías. Además, se pretende mostrar el proceso metodológico para construir, verifica y validar un MBA. El presente capítulo está compuesto por cinco apartados: el primer apartado muestra la evolución y las tendencias de investigación en el área de la MBA aplicada a los fenómenos de la CTI, especialmente en un marco de transferencia y difusión de tecnologías; el segundo apartado muestra los diferentes paradigmas de simulación y se centra en abordar el modelo metodológico y las herramientas computacionales para la construcción de un MBA. En el tercer apartado, se muestran las ventajas y desventajas que presenta el método frente a su aplicación, además, se resalta una de las características más relevantes de este, que es el apoyo a la construcción de teoría dado su enfoque de "abajo hacia arriba". En el último apartado se establecen las consideraciones para la aplicación del método especialmente en Latinoamérica y en los estudios de CTI.

Evolución de la investigación en modelación y simulación aplicados a la difusión y transferencia de tecnologías

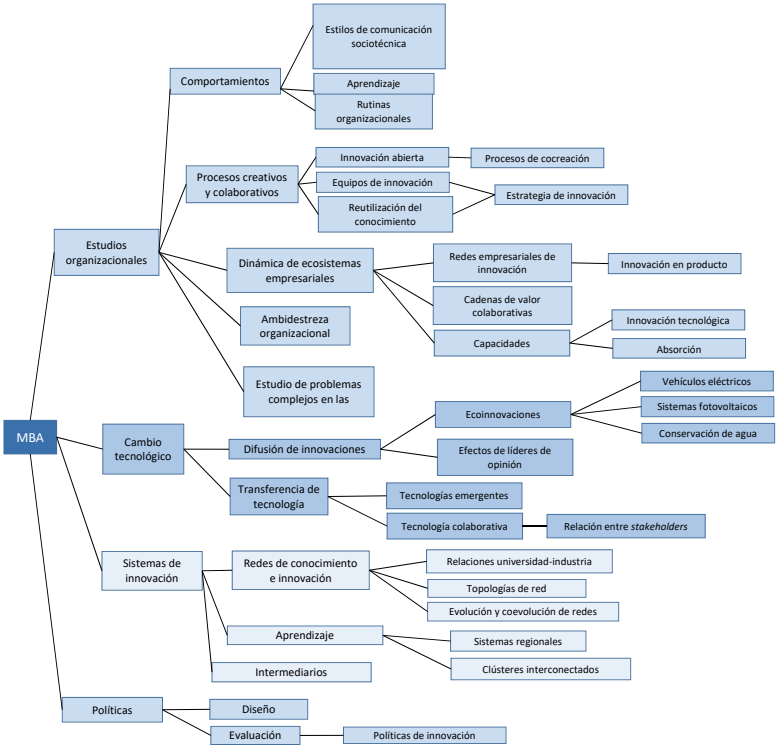
La literatura especializada muestra un amplio espectro de aplicaciones de la modelación basada en agentes para el estudio de los fenómenos de CTI, como se muestra en el diagrama 1, predominan cuatro grandes áreas de interés: estudios organizacionales, cambio tecnológico, sistemas de innovación y el diseño y evaluación de políticas. Respecto a los estudios organizacionales, la literatura muestra interés por el análisis de los comportamientos gerenciales, los procesos creativos colaborativos, las dinámicas de los ecosistemas empresariales y la ambidestreza organizacional.

En cuanto al estudio de los sistemas de innovación aplicando MBA, el interés se centra en el análisis de las redes de conocimiento e innovación, el aprendizaje que se da en este tipo de sistemas y el impacto de los intermediarios en los procesos de innovación y transferencia de tecnología. Para el área de política, la investigación se centra en el uso de la MBA para el diseño y evaluación de políticas de innovación o ecoinnovación.

Especialmente en el área de cambio tecnológico se presentan dos subáreas de interés, la primera relacionada con la difusión de innovación o tecnologías, especialmente la introducción de ecoinnovaciones como lo son los vehículos eléctricos, los sistemas fotovoltaicos residenciales o las tecnologías para la reducción del consumo de agua; de igual manera, se analiza el efecto que tienen los líderes de opinión sobre las decisiones y comportamientos de adopción. En cuanto a la transferencia de tecnología, se evidencia un interés en el análisis del efecto de tecnologías emergentes sobre el proceso de transferencia, dadas las características que estas presentan.

Dando una mirada mucho más profunda a la difusión y transferencia de tecnologías (en adelante, TT), la literatura científica muestra que estos paradigmas de simulación han sido empleados para mejorar la comprensión de los procesos de TT, a partir del desarrollo de modelos teóricos de adopción de tecnología, sobre la base del entendimiento de la heterogeneidad de los agentes (Bingham, Davis y Eisenhardt, 2007). En efecto, la MBA presenta una alternativa promisoria para intentar entender cómo los procesos sociales funcionan a través del tiempo (Ceschi *et al*., 2017), debido principalmente a que la MBA permite modelar los comportamientos e interacción de los agentes en el ambiente, empleando tanto datos cuantitativos como cualitativos para definir dichas relaciones (Filatova y Muelder, 2018). Otro aspecto interesante es que la MBA permite incluir factores estocásticos que generan una recreación más realista de los actos que se generan por arte de los agentes y que podrían no estar acordes con los patrones preestablecidos (Ceschi *et al*., 2017).

Diagrama 1. Áreas de investigación para la MBA aplicada a la CTI



Fuente: elaboración propia

Autores como Berger, Quang y Schreinemachers (2014) evaluaron cómo la interacción dinámica entre la pérdida de fertilidad del suelo y la toma de decisiones en granjas afecta el uso de métodos de conservación del suelo. Más específicamente, el estudio tiene como objetivo evaluar *ex ante* el potencial de tres métodos de bajo costo para la conservación de suelos implementando modelación basada en agentes.

Berger y Schreinemachers (2011) describen un software de simulación denominado sistema multiagentes basado en programación matemática (*mathematical programming-based multi-agent systems* (MP-MAS)), para el desarrollo de modelos basados en agentes, el cual está construido bajo el uso de una optimización con restricciones para simular la toma de decisiones en sistemas agrícolas. Su propósito es entender cómo la tecnología

agrícola, los mercados dinámicos, el cambio ambiental y la intervención política pueden afectar una población heterogénea de agricultores y los recursos agroecológicos que estos emplean.

Otra aplicación de la MBA para el proceso de TT ha sido desarrollada por Beretta *et al*. (2018). El estudio proporciona un modelo teórico de adopción de tecnología basado en la idea de que la difusión de información sobre una tecnología depende tanto de la estructura social de los adoptantes como del grado de asertividad.

Azad y Ghodsypour (2017) utilizan la modelación como mecanismo para identificar las relaciones entre agentes y mercado en un sistema híbrido denominado tecnosectorial, identificando las funciones de comportamiento dentro del ciclo de vida tecnológico en la industria petroquímica. Los resultados muestran que, para los dos escenarios definidos, la relación entre competidores y emprendedores juega un papel clave en la relación de beneficio en el sistema.

Otros modelos de simulación analizan los procesos de TT desde un enfoque de aprendizaje tecnológico (Giraldo Ramírez y Quintero Ramírez, 2018) para distintos sectores, como la base del cambio técnico y la transición tecnológica, en las que la simple existencia de la tecnología no implica la adopción y uso de la misma (Raven y Walrave, 2018).

Descripción del método y de herramientas para su aplicación

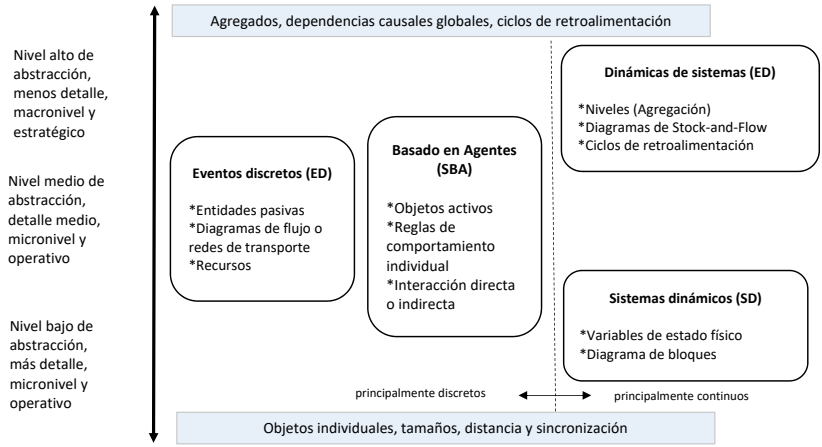
La modelación y simulación permiten resolver problemas de la vida real que no pueden ser prototipados o experimentados en condiciones reales debido a que es imposible o altamente costoso (Borshchev y Filippov, 2004). Los paradigmas o propuestas de modelación y simulación se han agrupado en cuatro bloques principales (ver figura 1), entre los que se encuentran: 1) dinámica de sistemas (DS), 2) simulación de eventos discretos (SED), 3) simulación basada en agentes (SBA) y 4) sistemas dinámicos (SD). DS y SED han sido ampliamente utilizados en la literatura; la SBA es un paradigma mucho más nuevo y los SD se emplean para modelar y diseñar sistemas físicos (Borshchev y Filippov).

Según el nivel de abstracción, los sistemas dinámicos se encuentran en la base de la figura 2; estos son usados en disciplinas de la ingeniería como mecánica, eléctrica, química y otras. Trata con variables continuas como velocidad, aceleración, presión y concentración. En un nivel medio de abstracción se encuentra la simulación de eventos discretos y la

simulación basada en agentes con un nivel medio de detalle. La primera es más usada en manufactura, logística o centros de servicio, mientras que la segunda se aplica a ciencias de la complejidad, inteligencia artificial y teoría de juegos (Borshchev y Filippov, 2004).

En un último nivel de abstracción superior, más global y de menos detalle se encuentra la dinámica de sistemas y simulación basada en agentes, cuyo foco se orienta al tratamiento de problemas de índole industrial, organizacional, dinámicas poblacionales y negociación, entre otros (Borshchev y Filippov, 2004). La MBA tiene un enfoque de modelado que se entiende como de "abajo hacia arriba"; el sistema no es controlado por un planificador central como un todo, sino que se controla en el nivel agregado el comportamiento de los agentes individuales (Happe, 2004).

Figura 1. Paradigmas de la modelación y simulación según la escala de abstracción



Fuente: adaptado de Borshchev y Filippov (2004)

Modelación basada en agentes (MBA)

Los modelos basados en agentes tienen su origen en tres campos de investigación diferentes: teoría de juegos, ciencia de la complejidad e inteligencia artificial distribuida (Elsenbroich y Gilbert, 2014). El modelado

y la simulación tienen el objetivo de reproducir e imitar el comportamiento de un sistema en la vida real (Guerrero, Schwarz y Slinger, 2016). El modelado basado en agentes es una técnica formada de las ciencias sociales computacionales que involucra la construcción de modelos que son programas de computador (Gilbert, 2007).

Esta técnica ha sido ampliamente usada en las ciencias sociales ya que involucra la creación de modelos que contemplan agentes, cada uno de los cuales representa un actor social y un ambiente en el cual estos actúan. Los agentes son capaces de interactuar entre ellos y son programados para ser proactivos, autónomos y capaces de percibir el mundo virtual (Gilbert, 2008). Los modelos computacionales son formulados como programas de computador en los cuales hay algunas entradas (como variables independientes) y algunas salidas (como variables dependientes). Los patrones, estructuras y comportamientos que emergen se manifiestan por la interacción entre agentes (Macal y North, 2010). Se pude entonces obtener a partir del modelo dos tipos de emergencia, emergencia imprevista e imprevisible. Un ejemplo de emergencia imprevista ocurre cuando se busca un estado de equilibrio, pero aparece algún tipo de comportamiento cíclico; los determinantes del comportamiento cíclico están ocultos en la estructura del modelo. El surgimiento impredecible ocurre cuando, por ejemplo, el caos aparece en los datos producidos por un experimento (Gilbert y Terna, 2000).

Una estructura básica de un modelo basado en agentes contiene tres elementos (Gilbert y Terna):

1. Un conjunto de agentes, sus atributos y comportamientos.
2. Un conjunto de relaciones de agentes y métodos de interacción: una topología de conexiones define cómo y con quién interactúan los agentes.
3. El entorno de los agentes: los agentes interactúan con su entorno además de otros agentes.

Agentes: los agentes son partes distintas de un programa que se utiliza para representar los actores sociales: personas, organizaciones (por ejemplo, empresas) o entidades (por ejemplo, Estados nacionales). Una característica crucial de los agentes es que pueden pasar mensajes entre ellos y actuar sobre la base de lo que aprenden de estos mensajes (Gilbert, 2007), de igual manera, los agentes están dotados de comportamientos que les permiten tomar decisiones independientes.

Los comportamientos de dichos agentes deben ser adaptativos y capaces de aprender de las experiencias que se derivan de la interacción y de las reglas de comportamiento, para ser llamados agentes, Macal y North (2010) definen las características esenciales de estos:

Cuadro 1. Características de los agentes

Característica	Definición
Autónomo, modular e individual	El agente tiene unas fronteras claras que permite fácilmente determinar si algo es parte o no del agente, el agente es diferenciado de otros, por lo que puede ser distinguido y reconocido
Auto dirigido	El agente es independiente en el ambiente y puede interactuar con otros según el modelo. El agente tiene comportamientos que se relacionan con la toma de decisiones
Presenta estados	El agente representa un estado que se relaciona con variables en una determinada situación. Los comportamientos de un agente están condicionados a su estado. Como tal, cuanto más rico sea el conjunto de estados posibles de un agente, más rico será el conjunto de comportamientos que puede tener un agente

Fuente: elaborado a partir de Macal y North (2010)

De igual manera, los agentes están dotados de comportamientos que les permiten tomar decisiones independientes. Las características que los condicionan son la percepción, actuación, memoria y políticas. Un agente puede percibir el ambiente y otros agentes en el entorno, exhibiendo comportamientos capaces de reproducir como comunicación e interacción. Los agentes presentan memoria, es decir, pueden recordar estados previos y acciones, y, por último, tienen reglas o estrategias que determinan su situación (Elsenbroich y Gilbert, 2014).

Interacción de los agentes: un MBA se refiere en sí mismo a la modelación de las relaciones e interacciones de los agentes y como estos se comportan y "toman decisiones" (Macal y North, 2010), en ese sentido, el agente específicamente conoce quién es, qué puede hacer o con quién puede conectarse. Esta información característica de cada agente le ofrece al modelo su característica descentralizada, es decir, no existe una autoridad que controle todos los agentes, sino que por el contrario, los agentes controlan su comportamiento en un esfuerzo de optimizar el rendimiento del sistema (Macal y North).

Un agente interactúa con agentes que están en su vecindario, pero a medida que la simulación se desarrolla estos pueden cambiar o moverse

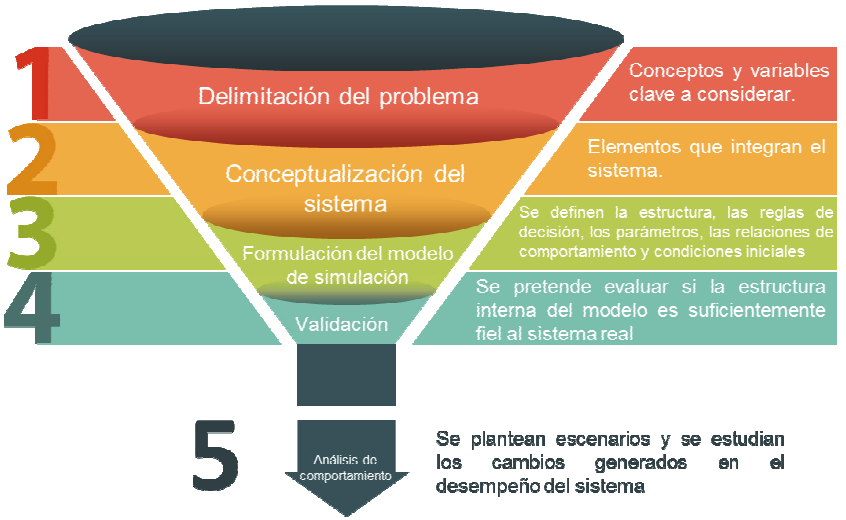
en el ambiente. Generalmente, los agentes están conectados por una topología, esta establece quién transfiere información a quién, por lo que según el modelo o fenómeno a estudiar se deberá establecer la tipología, entre estas se encuentran: redes, espacios euclidianos, autómata celular, sistemas de información geográfica o modelos espaciales (Macal y North).

Entorno o ambiente de los agentes: el entorno es el mundo virtual en el que actúan los agentes, comúnmente, los entornos representan espacios geográficos (Gilbert, 2007). Estos proveen información de la localización espacial del agente, por lo que el ambiente podría establecer restricciones para la interacción entre agentes (Macal y North)

Procedimientos/etapas para su aplicación

La metodología planteada para la construcción de un modelo basado en agentes contempla cinco fases, sustentadas en una adaptación del modelo para la construcción y la validación de un modelo de simulación de Sargent (2013) (ver figura 2).

Figura 2. Metodología para la modelación y simulación



Fuente: elaboración de los autores a partir de Sargent (2013) y Quintero (2016)

Fase 1: delimitación del problema

El primer paso para desarrollar un proceso de modelación es establecer claramente cuál es el problema de investigación, por qué este es un problema y cuáles deben ser los conceptos, variables clave y comportamientos problemáticos que se deberán considerar. Es importante ser claro, conciso y específico para la formulación del modelo, es decir, informar por qué es necesario construirlo y qué se espera obtener con este (Bastiansen *et al*., 2006).

Fase 2: conceptualización del sistema

Se definen los elementos que integran la descripción del sistema, así como las influencias que se producen entre estos. Asimismo, se establece la frontera del sistema, es decir, qué es considerado endógeno o exógeno.

Para esta fase puede aplicarse el proceso de construcción propuesto por Wilensky (1999), que plantea tres etapas y que posteriormente servirá de fundamento para el desarrollo de la MBA. La primera etapa consiste en realizar unas preguntas iniciales que permiten dar claridad sobre cómo aporta el modelo a la comprensión del fenómeno, entre estas se encuentran:

- ¿Cuál es la pregunta de investigación?
- ¿Qué se quiere modelar?
- ¿Qué ideas requieren ser examinadas?
- ¿Cuáles son los detalles esenciales del sistema y cuáles no?
- ¿Cómo ayuda el modelo a la comprensión del fenómeno?

La segunda etapa consiste en trasladar las respuestas obtenidas a la teoría, buscando que el modelo refleje de forma adecuada los conceptos que lo soportan teóricamente. En la tercera etapa se formulan hipótesis que permitan construir el modelo conceptual.

Fase 3: formulación del modelo de simulación

En este paso se definen la estructura, las reglas de decisión, los parámetros, las relaciones de comportamiento y las condiciones iniciales. Este debe ser verificado computacionalmente antes de que pueda ser utilizado, para evaluar la consistencia y el cumplimiento de la delimitación propuesta.

Supuestos del modelo: son las premisas bajo las cuales se construye el modelo, es decir, los argumentos que soportan el funcionamiento de este.

Modelo conceptual: representa los procedimientos y los vínculos que se generan entre los agentes en el entorno, además de las clases generales de agentes, su cantidad aproximada (Rand y Rust, 2011), propiedades y comportamientos que los describen.

Reglas de decisión: las interacciones entre los agentes se dan a través de mecanismos o reglas de decisión, los cuales pueden ser descriptos usando un diagrama de flujo del modelo. A través del diagrama de flujo se busca también aclarar el orden en el que se dan los procesos, intentando dar respuesta a cómo pasan realmente las acciones y qué acciones son fijas, sincrónicas o asincrónicas.

Parámetros y parametrización: se debe aclarar qué procesos, individuales o del ambiente, se construirán en el modelo (Bastiansen *et al.*, 2006), para llevar a cabo las simulaciones, y cuáles son los valores elegidos de cada uno de los parámetros. Se debe establecer si estos son seleccionados arbitrariamente o basados en datos.

Construcción del modelo de simulación: el objetivo es crear e implementar una versión del modelo que pueda ser ejecutada computacionalmente y que corresponda con el modelo conceptual antes planteado (Rand y Rust). Esta programación consiste en la construcción de varios procedimientos, equivalentes a submodelos que componen todo el modelo.

Verificación computacional: la verificación determina el nivel de correspondencia entre el modelo implementado y el modelo conceptual, es decir, asegura que el programa computacional y la implementación del modelo conceptual son los correctos (Sargent, 2013). Se aplicará la técnica de validación de trazas que consiste en el seguimiento de los comportamientos de las entidades en cada submodelo y del modelo general, para determinar si la lógica es correcta y si es obtenida la precisión necesaria.

Fase 4: validación del modelo

Debido a que todos los modelos, mentales o formales, son representaciones limitadas y simplificadas del mundo real, estos difieren de la realidad. Este paso pretende evaluar si la estructura interna del modelo es suficientemente fiel al sistema real. Se ponen a prueba los supuestos y reglas de decisión establecidas, así como el comportamiento del modelo.

Se realizará una validación conceptual y otra operacional. La primera busca garantizar que el modelo está correctamente sustentando según los supuestos y reglas de decisión; la segunda, la validación operacional, busca que el comportamiento resultante capture la dinámica real del sistema, con un rango de precisión satisfactorio. Sargent describe una serie de técnicas y test usados para la verificación y validación de modelos de simulación, entre las que se encuentran: animación, comparación con otros modelos, pruebas degeneradas, validez de eventos, pruebas en condiciones extremas, validez de cara, validación histórica de datos, métodos históricos (racionalismo-empirismo), trazas y validez multietapas, entre otras técnicas.

De igual manera, otras propuestas de validación son desarrolladas por Fagiolo, Moneta y Windrum (2007) como la calibración indirecta y el modelo histórico amigable (*history friendly models* (HFM)). Este último busca llevar el modelo más cerca de la evidencia empírica. La diferencia de este enfoque es que utiliza casos históricos específicos de una industria o sector para obtener los parámetros del modelo, interacciones y reglas de decisión de los agentes (Fagiolo, Moneta y Windrum). Este enfoque ha sido desarrollado en diferentes estudios (Malerba *et al.*, 1999). Para la validación conceptual se empleará el método histórico del racionalismo y para la validación operacional se empleará el modelo histórico amigable.

Fase 5: análisis del comportamiento del modelo

Se aplica un análisis de sensibilidad para mejorar el entendimiento del comportamiento del modelo según los parámetros, condiciones iniciales, delimitación del modelo y agregación, esto permitirá identificar parámetros sensibles. En esta etapa es definida una base para la comparación de los comportamientos. Posteriormente, se plantean escenarios y se estudian los cambios generados, comparando los diferentes resultados para analizar las dinámicas de interacción y los

patrones de comportamiento en los procesos de TT. Finalmente, se evalúa un conjunto de políticas para valorar el efecto de estas sobre el sistema de estudio.

Finalmente, los resultados obtenidos con el modelo computacional deben permitir la validación, a partir de la comparación con datos empíricos observables sobre el fenómeno de estudio, es decir, el modelo representa con claridad y satisfacción el fenómeno. Logrando capturar los elementos fundamentales para el análisis de su desempeño. De este modo se pueden determinar, según el rigor en la construcción del modelo conceptual su verificación y validación, distintos niveles de desempeño del modelo (Axtell y Epstein, 1994):

- Nivel 0: el modelo es una "caricatura de la realidad", tal como se establece mediante el uso de dispositivos gráficos simples (por ejemplo, que permite simplemente la visualización del movimiento del agente).
- Nivel 1: el modelo está en acuerdo cualitativo con las macroestructuras empíricas, según las distribuciones de algunos atributos de la población de agentes.
- Nivel 2: el modelo produce un acuerdo cuantitativo con las macroestructuras empíricas, según lo establecido a través de rutinas de estimación estadística.
- Nivel 3: el modelo exhibe un acuerdo cuantitativo con las microestructuras empíricas, según lo determinado a partir del análisis transversal y longitudinal de la población de agentes.

Herramientas de aplicación

Existen diversas herramientas para la implementación de modelos de simulación basados en agentes, entre las que se encuentran NetLogo y AnyLogic.[2] La primera es una interfaz de programación para MBA, que pertenece al enfoque de modelación de "abajo hacia arriba" con un nivel de abstracción medio, sin embargo, la segunda herramienta permite combinar enfoques tanto de "abajo hacia arriba", propio de la

2 Cabe destacar que existen otras herramientas como Vensim, Stella, Modelica, entre otras.

MBA, como de "arriba hacia abajo", propio de la dinámica de sistemas. Esta herramienta permite combinar niveles de abstracción superior y medio en la representación de femémonos de la realidad para la construcción de modelos.

NetLogo[3] es un entorno de modelado programable multiagente, fue desarrollado por Uri Wilensky en 1999, fundador y actual director del Centro de Aprendizaje Conectado y Modelado Computarizado (CCL, por sus siglas en inglés) y cofundador del Instituto Noroeste de Sistemas Complejos (NICO). Es un software gratuito de escritorio con una versión de uso en la web, y se orienta principalmente a modelar sistemas complejos que se desarrollan en el tiempo. Es posible programar miles de agentes que operan independientemente, lo que permite observar la emergencia de patrones en el macronivel a partir de comportamientos en el micronivel de los individuos. NetLogo permite también explorar las condiciones de simulación bajo diferentes escenarios.

AnyLogic[4] es un software de simulación pago para empresas e industrias, desarrollado por The AnyLogic Company en el año 2000. La simulación permite experimentar con una representación digital válida de un sistema, inspeccionar procesos e interactuar con un modelo en acción, mejorando la comprensión de este. Permite vincular simultáneamente paradigmas de "abajo hacia arriba" y de "arriba hacia abajo" (modelos híbridos).

Datos

El tipo de datos usados para la simulación basada en agentes difiere de los datos usados para otros tipos de simulación. Específicamente, la MBA puede usar datos cuantitativos, etnográficos o cualitativos, esto depende del fenómeno que se esté desarrollando y los elementos conceptuales para su construcción y validación (Elsenbroich y Gilbert, 2014).

3 Más información, ver: https://ccl.northwestern.edu/netlogo/.
4 Más información, ver: https://www.anylogic.com/.

Ventajas y desventajas de la MBA

Entre las ventajas de los modelos de simulación basados en agentes se encuentra la aproximación dinámica a los fenómenos sociales, en la que principalmente se realizan estudios basados en teorías y atemporales. Otro elemento a destacar es la heterogeneidad de los agentes, debido a que principalmente los modelos matemáticos tienden a asumir que los agentes son uniformes para facilitar los procesos de cálculo. En muchos modelos económicos basados en teoría económica se hacen suposiciones simplificadoras para la trazabilidad analítica. Estos supuestos incluyen 1) los agentes económicos racionales, lo que implica que los agentes tienen objetivos bien definidos y son capaces de optimizar su comportamiento, 2) los agentes económicos son homogéneos, es decir, los agentes tienen características y reglas de comportamiento idénticas, 3) el sistema experimenta principalmente rendimientos decrecientes a escala de los procesos económicos (disminución de la utilidad marginal, disminución de la productividad marginal, etc.) y 4) el estado de equilibrio a largo plazo del sistema es la información principal de interés (Macal y North, 2010).

Un tercer elemento de interés para los MBA es que se pueden modelar tanto microcomportamientos como macroinfluencias de esos microcomportamientos, mediante ciclos de retroalimentación que le dan un enfoque dinámico a los modelos (Elsenbroich y Gilbert, 2014; Ceschi *et al*., 2018). En otras palabras, las decisiones de los agentes tienen un efecto general en el comportamiento de las estructuras sociales superiores, que a su vez afecta las decisiones posteriores de los agentes individuales en el siguiente paso del modelo (Ceschi *et al*.).

Un cuarto elemento a destacar de la MBA es que es una herramienta metodológica que se puede utilizar para aprovechar al máximo los datos escasos y desiguales (Gilbert y Terna, 2000). Al usar relativamente pocos datos (tanto cualitativos como cuantitativos y etnográficos), se puede generar reglas de comportamiento para los agentes, con la hipótesis de que lo que se encuentra a pequeña escala es generalizable. Al ejecutar la simulación, podemos probar esta hipótesis comparando los resultados de la simulación con los datos del mundo real (Elsenbroich y Gilbert).

Uno de los principales problemas asociados a la MBA se relaciona directamente con la validación empírica, esto debido principalmente a la relación entre los modelos y el "sistema real" que se está modelando

(Fagiolo, Moneta y Windrum, 2007), es decir, la validación involucra la evaluación del grado en el cual el modelo representa un fenómeno real. Por lo general, la validación se realiza comparando los resultados de la simulación con datos empíricos del sistema objetivo. A menudo, estos datos son de naturaleza estadística, pero para algunos modelos muy abstractos, la validación solo puede ser posible mediante una comparación cualitativa o incluso solo una inspección visual (Elsenbroich y Gilbert). Los modelos simples, con menor número de parámetros y procesos de representación sencillos, son más fáciles de calibrar y validar, en contraposición con los modelos complejos, con múltiples submodelos interconectados (Balbi *et al*., 2016).

Otro elemento de amplia discusión en la literatura de la MBA es la falta de comparabilidad entre modelos que han sido desarrollados. Los modelos no solo tienen un contenido teórico diferente que los sustente, sino que buscan explicar fenómenos similares, que en el momento de comparar y evaluar el poder explicativo de cada uno no resulta tan sencillo, en ocasiones los modelos altamente complejos y difíciles de comprender son denominados como "cajas negras" (Hansen *et al*., 2003). Ante los problemas asociados a la validación y la comparabilidad de los modelos, Edmonds y Moss (2005) plantean un enfoque *keep it descriptive stupid* (KIDS), es decir, el modelo debe construirse lo suficientemente complicado y detallado como para modelar la riqueza de los sistemas; también reconocen que la mezcla de los enfoques KISS (*keep it simple stupid*) y KIDS puede probablemente ser más apropiada (Balbi *et al*., 2016).

Protocolo ODD

Con el objetivo de hacer frente a las principales desventajas de la MBA, algunos autores han definido estructuras de modelado que permitan la replicabilidad, la trazabilidad y la comparación entre fenómenos analizados a partir de la abstracción de elementos principales que deberán contemplarse en cualquier modelo para evitar la ambigüedad.

El modelo consiste en tres bloques (Resumen, Conceptos de diseño y Detalles), que se subdividen en siete elementos: Propósito, Variables y escalas de estado, Resumen y programación del proceso, Conceptos de diseño, Inicialización, Entrada y Submodelos (Bastiansen *et al*., 2006).

Cuadro 2. Estructura del modelado

Bloques	Elementos	Descripción
Resumen	Propósito	Debe ser declarado inicialmente qué se espera obtener con el modelo y cuál es el objetivo para su construcción.
	Variables y escalas de estado	Se deben definir claramente cuáles son los agentes que componen el modelo, qué tipo de jerarquías se presenta, y también las variables que lo caracteriza.
	Resumen y programación del proceso	Se deben definir los procesos que ocurren en el modelo, todos, y de manera detallada. Además, es importante aclarar cómo ocurren a través del tiempo para poder determinar si usar tiempo continuo o discreto.
Conceptos de diseño	Conceptos de diseño	Se refiere a un checklist de elementos que deberán estar claros en el momento de plantear el modelo, entre estos están: emergencia, adaptación, predicción, interacción, sentido, colectividad y observación. Para una descripción ampliada ver: Bastiansen *et al.* , 2006.
Detalles	Inicialización	Se debe definir cómo se crean los individuos y el ambiente del sistema cuando la modelación inicia. Los valores iniciales de la simulación son aleatorios o basados en datos.
	Entrada	Para definir la predictibilidad del modelo, se debe establecer qué datos se cargan inicialmente al sistema y cuáles otros son aleatorios.
	Submodelos	Las variables presentadas en el apartado de "Variables y escalas de estado" deberán ser explicadas en detalle, estableciendo cuáles son las reglas y los modelos matemáticos que sustentan el modelo.

Fuente: elaboración propia con base en Bastiansen *et al.* (2006)

El protocolo ODD ha sido ampliamente implementado en la construcción de diferentes modelos basados en agentes, y actualmente muchos de los artículos científicos realizan la descripción bajo los siete elementos que lo componen, mejorando la replicabilidad y reduciendo las desventajas de comparabilidad entre modelos.

Construcción de teoría y toma de decisiones mediante MBA

Uno de los aspectos más relevantes de la aplicación de la MBA es su potencial aporte en la creación de teoría (Bingham, Davis y Eisenhardt, 2007), esto debido a que facilita la compresión y confirmación de relaciones teóricas complejas entre constructos (Epstein, 2008), especialmente cuando existen limitantes en los datos (Zott, 2003). Los MBA pueden apoyar la construcción de teorías combinando enfoques inductivos y deductivos en una llamada "tercera forma de hacer ciencia" (Axelrod, 1997). La MBA también puede ayudar a los tomadores de decisión a facilitar este proceso, principalmente por la facilidad que provee la herramienta para el análisis de escenarios (Balbi *et al*., 2016; Quintero, 2016). Los modelos robustos pueden apoyar la toma de decisiones, ya que se orientan a analizar fenómenos específicos de interés para un grupo de involucrados (Murray-Rust, Robinson y Rounsevell, 2012).

Aplicaciones en estudios de CTI y sociedad

El desarrollo teórico y conceptual de la TT en los CTI, y en particular, el de las cadenas productivas agropecuarias inmersas en estos tipos de sistemas, ha sido influenciado por diferentes escuelas de pensamiento; sin embargo, las contribuciones en la comprensión del fenómeno evolutivo de la TT bajo el enfoque de la teoría de la complejidad han tenido una mejor aceptación debido a que se permite el análisis de aquellos fenómenos denominados emergentes, como lo es la TT. Además, el análisis de estos sistemas debe ser de carácter dinámico, para obtener así una mejor aproximación a las dinámicas y los patrones que adoptan estos tipos de sistemas. Una metodología de análisis para hacer frente a este problema es la de la MBA, que se considera adecuada para abordar los *sistemas complejos adaptativos* que se desarrollan y evolucionan en el tiempo, como lo son las cadenas productivas agropecuarias. No obstante, la revisión de la literatura revela un esfuerzo apenas incipiente en el desarrollo de modelos conceptuales construidos para analizar el fenómeno de la TT a la luz de los MBA.

Consideraciones para la aplicación de la MBA en América Latina y el Caribe (LAC)

Como se mencionó antes, son múltiples las aplicaciones de la MBA orientadas a mejorar el entendimiento de fenómenos de CTI que se desarrollan bajo sistemas complejos sociales que evolucionan. De igual manera, son extensas las aplicaciones en el área de las ciencias sociales de las que toma su origen la MBA. Sin embargo, para el caso de LAC esta perspectiva de abordaje de problemas desde abajo hacia arriba ("bottom up"), apenas viene siendo implementada, es por esto que herramientas como la MBA cobran vital importancia, dado las problemáticas complejas que han persistido por años en los territorios de LAC, aceleradas por la pobreza, desigualdad, aumento de la criminalidad, cobertura de servicios públicos, desempleo y cambio climático (explotación de la riqueza natural), sin dejar de lado el clima político poco favorable.

En este sentido, la MBA puede servir para representar y comprender de mejor manera las complejas interacciones espaciales, bajo condiciones heterogéneas y procesos de toma de decisión de agentes que se adaptan, y que se suscriben en los países de LAC. La MBA permite realizar experimentos virtuales que pueden evidenciar la emergencia de patrones sociales, tales como la emergencia de normas, formas de coordinación y participación en acciones colectivas, bajo distintas topologías de red, modelos de movilidad espacial y estratificación social, entre otros.

Al realizar una revisión del uso y aplicación de esta herramienta en casos específicos para LAC, se puede observar que el enfoque está enmarcado desde la salud pública, en la que la modelación y simulación de dinámicas de propagación de vectores asociados a enfermedades epidémicas, como la enfermedad de Chagas (Devillers, Lobry y Menu, 2008; Cordovez y Erazo, 2013), la teniasis (Garcia *et al*., 2020), la chikunguña y el zika (Burke *et al*., 2018), son el centro del análisis para la búsqueda de estrategias de control, y más aún cuando se trata de enfermedades desatendidas y causantes de miles de muertes anuales.

De igual manera, el análisis de otro tipo de enfermedades, como se hizo con la enfermedad por el covid-19, aplicando AMB permite la construcción y análisis de, por ejemplo, políticas de mitigación, distanciamiento social, vacunación, rastreo de casos, aislamiento, control y gestión a largo plazo, permitiendo mejorar la comprensión de los efectos económicos del distanciamiento en los sistemas sociotécnicos. Este tipo de aplicaciones a problemas de nuestra realidad cercana muestra el potencial de uso de

la MBA para el diseño de estrategias, análisis de escenarios y toma de decisiones en diferentes campos del conocimiento, generando un efecto de masificación de su uso, dada su utilidad. Muestra de ello es que solamente en el año 2020 se han generado más de sesenta documentos científicos en la base de datos Scopus[5] para temáticas relacionadas con el SARS-CoV-2.

Conclusiones

La MBA ha tomado importancia en los últimos años, especialmente cuando se trata de tener en consideración las interacciones y los comportamientos de agentes heterogéneos, y la emergencia de los patrones que se deriven de estas interacciones. Los métodos basados en agentes se han aplicado en este contexto tanto en la construcción de teorías como en las herramientas para analizar escenarios del mundo real, apoyar decisiones de gestión y obtener recomendaciones de políticas (Günther *et al*., 2012).

El modelo metodológico presentado en este capítulo integra todos los elementos a tener en cuenta para facilitar la construcción de una MBA, además, de las diferentes herramientas para su validación y verificación. Por lo anterior, la MBA puede convertirse en una herramienta de interés y relevancia para el análisis de los fenómenos de CTI en Latinoamérica, pues permitirá programa en un modelo virtual las características homogéneas que nos gobiernan como región.

Bibliografía

Axelrod, R. (1997). "Advancing the Art of Simulation in the Social Sciences". En Conte, R.; Hegselmann, R. y Terna, P. (eds.), *Simulating Social Phenomena. Lecture Notes in Economics and Mathematical Systems*, vol. 456. Berlín, Heidelberg: Springer. DOI: https://doi.org/10.1007/978-3-662-03366-1_2.

Axtell, R. y Epstein, J. (1994). "Agent-Based Modeling: Understanding Our Creations". *The Bulletin of the Santa Fe Institute*, vol. 9, nº 4, pp. 28-32.

5 Base de datos consultada en diciembre de 2020, bajo el siguiente algoritmo: ((TITLE-ABS-KEY (("agent-based model*") OR ("ABM")))) AND ("covid-19").

Azad, S. M. y Ghodsypour, S. H. (2017). "Modeling the dynamics of technological innovation system in the oil and gas sector". *Kybernetes*, vol. 47, n° 4, pp. 771-800. DOI: https://doi.org/10.1108/K-03-2017-0083.

Balbi, S.; Buchmann, C. M.; Groeneveld, J.; Lauf, S.; Lorscheid, I.; Magliocca, N. R.; Millington, J. D.; Müller, B.; Nolzen, H.; Schulze, J. y Sun, Z. (2016). "Simple or complicated agent-based models? A complicated issue". *Environmental Modelling and Software*, vol. 86, n° 56-67. DOI: https://doi.org/10.1016/j.envsoft.2016.09.006.

Bastiansen, F.; Berger, U.; DeAngelis, D. L.; Eliassen, S.; Ginot, V.; Giske, J.; Goss-Custard, J.; Grand, T.; Grimm, V.; Heinz, S. K.; Huse, G.; Huth, A.; Jepsen, J. U.; Jørgensen, C.; Mooij, W. M.; Müller, B.; Pe'er, G.; Piou, C.; Railsback, S. F.; Robbins, A. M.; Robbins, M. M.; Rossmanith, E.; Rüger, N.; Souissi, S.; Stillman, R. A.; Strand, E.; Vabø, R. y Visser, U. (2006). "A standard protocol for describing individual-based and agent-based models". *Ecological Modelling*, vol. 198, n° 1-2, pp. 115-126. DOI: https://doi.org/10.1016/j.ecolmodel.2006.04.023.

Beretta, E.; Fontana, M.; Guerzoni, M. y Jordan, A. (2018). "Cultural dissimilarity: Boon or bane for technology diffusion?". *Technological Forecasting and Social Change*, vol. 133, pp. 95-103. DOI: https://doi.org/10.1016/j.techfore.2018.03.008.

Berger, T. y Schreinemachers, P. (2011). "An agent-based simulation model of human – environment interactions in agricultural systems". *Environmental Modelling & Software*, vol. 26, n° 7, pp. 845-859. DOI: https://doi.org/10.1016/j.envsoft.2011.02.004.

Berger, T.; Quang, D. V. y Schreinemachers, P. (2014). "*Ex-ante* assessment of soil conservation methods in the uplands of Vietnam: An agent-based modeling approach". *Agricultural Systems*, vol. 123, pp. 108-119. DOI: https://doi.org/10.1016/j.agsy.2013.10.002.

Bingham, C. B.; Davis, J. P. y Eisenhardt, K. M. (2007). "Developing Theory Through Simulation Methods". *Academy of Management Review*, vol. 32, n° 2, pp. 480-499. DOI: https://doi.org/10.5465/amr.2007.24351453.

Borshchev, A. y Filippov, A. (2004). "From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques,

Tools". Presentado en el *22nd International Conference of the System Dynamics Society*, 25-29 de julio. Oxford, Inglaterra.

Burke, D. S.; Campo Carey, A.; de la Hoz, F.; Diaz, H.; España, G.; Grefenstette, J.; Perkins, A.; Torres, C. y van Panhuis, W. G.(2018). "Exploring scenarios of chikungunya mitigation with a data-driven agent-based model of the 2014-2016 outbreak in Colombia". *Scientific Reports*, vol. 8, n°1. DOI: https://doi.org/10.1038/s41598-018-30647-8.

Carlsson, B. y Stankiewicz, R. (1991). "On the nature, function and composition of technological systems". *Journal of Evolutionary Economics*, vol. 1, n° 2, pp. 93-118. DOI: https://doi.org/10.1007/BF01224915.

Ceschi, A.; Dickert, S.; Frayret, J.; Sartori, R.; Scalco, A. y Shiboub, I. (2017). *Agent-Based Modeling of Sustainable Behaviors*, pp. 77-97. DOI: https://doi.org/10.1007/978-3-319-46331-5. Cham: Springer.

Ceschi, A.; Sartori, R. y Scalco, A. (2018). "Application of Psychological Theories in Agent-Based Modeling: The Case of the Theory of Planned Behavior". *Nonlinear Dynamics, Psychology, and Life Sciences*, vol. 22, n° 1, pp. 15-33.

Cordovez, J. M. y Erazo, D. C. (2013). "A spatial agent based model for the simulation of house infestation by R. prolixus the principal vector of Chagas disease in Colombia". Presentado en el *2013 Pan American Health Care Exchanges (PAHCE)*. Medellín, Colombia.

Devillers, H.; Lobry, J. R. y Menu, F. (2008). "An agent-based model for predicting the prevalence of *Trypanosoma cruzi* I and II in their host and vector populations". *Journal of Theoretical Biology*, vol. 255, n° 3, pp. 307-315. DOI: https://doi.org/10.1016/j.jtbi.2008.08.023.

Edmonds, B. y Moss, S. (2005). "From KISS to KIDS – An 'Anti-simplistic' Modelling Approach". En Davidsson, P.; Logan, B. y Takadama, K. (eds.), *Multi-Agent and Multi-Agent-Based Simulation*, pp. 130-144. Heidelberg, Berlín: Springer/Verlag. DOI: https://doi.org/10.1007/978-3-540-32243-6_11.

Elsenbroich, C. y Gilbert, N. (2014). "Modelling Norms". *Modelling Norms*, pp. 143-149. Países Bajos: Springer. DOI: https://doi.org/10.1007/978-94-007-7052-2.

Epstein, J. (2008). "Why model?". *Journal of Artificial Societies and Social Simulation*, vol., 11, n° 4. Disponible en: http://jasss.soc.surrey.ac.uk/11/4/12.html.

Fagiolo, G.; Moneta, A. y Windrum, P. (2007). "Empirical validation of agent-based models: Alternatives and prospects". *Journal of Artificial Societies and Social Simulation*, vol. 10, n° 2, pp. 8-38. Disponible en: http://jasss.soc.surrey.ac.uk/10/2/8.html.

Filatova, T. y Muelder, H. (2018). "One theory-many formalizations: Testing different code implementations of the theory of planned behaviour in energy agent-based models". *Journal of Artificial Societies and Social Simulation*, vol. 21, n° 4. DOI: https://doi.org/10.18564/jasss.3855.

Garcia, H. H.; Gonzalez, A. E.; Lambert, W. E.; O'Neal, S. E.; Pan, W.; Pray, I. W. y Wakeland, W. (2020). "Understanding transmission and control of the pork tapeworm with CystiAgent: a spatially explicit agent-based model". *Parasites & Vectors*, vol. 13, n° 1. DOI: https://doi.org/10.1186/s13071-020-04226-8.

Gilbert, N. (2007). *Agent-Based Models*. Thousand Oaks, CA: SAGE.

_____ (2008). *Agent Based Models*. Thousand Oaks, CA: SAGE. DOI: https://doi.org/10.4135/9781412983259.

Gilbert, N. y Terna, P. (2000). "How to build and use agent-based models in social science". *Mind & Society*, vol. 1, n° 1, pp. 57-72. DOI: https://doi.org/10.1007/BF02512229.

Giraldo Ramírez, D. y Quintero Ramírez, S. (2018). *El aprendizaje en los sistemas regionales de innovación desde la perspectiva de la modelación basada en agentes*. Colombia: Universidad Pontificia Bolivariana.

Guerrero, C. N.; Schwarz, P. y Slinger, J. (2016). "A recent overview of the integration of System Dynamics and Agent-based Modelling and Simulation". *34th International Conference of the System Dynamics Society*. Delft, Países Bajos.

Günther, M.; Kiesling, E.; Stummer, C. y Wakolbinger, L. M. (2012). "Agent-based simulation of innovation diffusion: a review". *Central European Journal of Operations Research*, vol. 20, n° 2, pp. 183-230. DOI: https://doi.org/10.1007/s10100-011-0210-y.

Hansen, T. S.; Jensen, T. S.; Jepsen, J. U.; Nikolajsen, F.; Odderskær, P. y Topping, C. J. (2003). "ALMaSS, an agent-based model for animals in temperate European landscapes". *Ecological Modelling*, vol. 167, n° 1-2, pp. 65-82. DOI: https://doi.org/10.1016/S0304-3800(03)00173-X.

Happe, K. (2004). *Agricultural policies and farm structures Agent-based modelling and application to EU-policy reform*. Halle: Institute of Agricultural Development in Central and Eastern Europe (IAMO).

Macal, C. M. y North, M. J. (2010). "Tutorial on agent-based modelling and simulation". *Journal of Simulation*, vol. 4, n° 3, pp. 151-162. DOI: https://doi.org/10.1057/jos.2010.3.

Malerba, F.; Nelson, R.; Orsenigo, L. y Winter, S. (1999). "'History-friendly' models of industry evolution: the computer industry". *Industrial and Corporate Change*, vol. 8, n° 1, pp. 3-40. DOI: https://doi.org/10.1093/icc/8.1.3.

Murray-Rust, D.; Robinson, D. T. y Rounsevell, M. D. A. (2012). "From actors to agents in socio-ecological systems models". *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 367, n° 1586, pp. 259-269. DOI: https://doi.org/10.1098/rstb.2011.0187.

Quintero, S. (2016). *Aprendizaje en los sistemas regionales de innovación: Un modelo basado en agentes*. Colombia: Universidad Nacional de Colombia.

Rand, W. y Rust, R. T. (2011). "Agent-based modeling in marketing: Guidelines for rigor". *International Journal of Research in Marketing*, vol. 28, n° 3, pp. 181-193. DOI: https://doi.org/10.1016/j.ijresmar.2011.04.002.

Raven, R. y Walrave, B. (2018). Overcoming transformational failures through policy mixes in the dynamics of technological innovation systems. *Technological Forecasting and Social Change*, vol. 153. DOI: https://doi.org/10.1016/j.techfore.2018.05.008.

Sargent, R. G. (2013). "Verification and validation of simulation models". *Journal of Simulation*, vol. 7, n° 1, pp. 12-24. DOI: https://doi.org/10.1057/jos.2012.20.

Tesfatsion, L. (2002). "Agent-Based Computational Economics: Growing Economies From the Bottom Up". *Artificial Life*, vol. 8, nº 1, pp. 55-82. DOI: https://doi.org/10.1162/106454602753694765

Wilensky, U. (1999). *NetLogo*. Evanston, IL: Center for Connected Learning and Computer Based Modeling/Northwestern University. Disponible en: http://ccl.northwestern.edu/netlogo/.

Zott, C. (2003). "Dynamic capabilities and the emergence of intraindustry differential firm performance: insights from a simulation study". *Strategic Management Journal*, vol. 24, nº 2, pp. 97-125. DOI: https://doi.org/10.1002/smj.288.

Capítulo 7
El análisis semántico-estadístico como estrategia de abordaje metodológico: reflexiones sobre su pertinencia en el estudio de problemáticas latinoamericanas

Rodrigo Kataishi, Matías Milia

Introducción

Durante el último cuarto de siglo, los procesos de digitalización de la investigación científica han sido uno de los principales ejes de transformación en las formas de producción de conocimiento. Esta dinámica ha sido aprovechada de forma inicial por disciplinas tradicionales relacionadas con las ciencias experimentales, como la biología o la física. En las ciencias sociales ha requerido algo más de tiempo (Adamic *et al*., 2009), dado que la naturaleza de los abordajes difiere en complejidad, estrategias y disponibilidad de información.

La digitalización de registros públicos, del trabajo profesional y de la vida social ha implicado un aumento en la cantidad de información disponible: hay datos producidos por sensores digitales en espacios físicos, otros que emergen del uso e interacción con plataformas digitales (Golder y Macy, 2014), pero también los hay generados por gobiernos, instituciones oficiales y profesionales dentro de nuestra región, involucrando también lo producido en los respectivos sistemas de ciencia y tecnología. La noción de que la digitalización de grandes cantidades de información ha permitido "hacer cuantificable lo incuantificable" (Watts, 2011) ha dado luz a varios programas de investigación en campos emergentes

como las ciencias sociales computacionales (Salganik, 2018), las humanidades digitales (Berry, 2012) o los estudios a partir de la utilización de los denominados métodos digitales (Rogers, 2013). Estas interpretaciones, que apuntan a distintas apropiaciones de "lo digital", tienden a otorgar relevancia a los procesos de diseminación y socialización de fuentes de información, y simultáneamente movilizan reflexiones que dialogan con diversas referencias teóricas (Castro Gouveia y Texeira Rabello, 2019: 145-147). Plantean, así, una novedad en la investigación social.

En dicho contexto, este capítulo busca realizar aportes que permitan capturar el potencial emergente para el estudio de los sistemas de ciencia, tecnología, innovación y sociedad (CTIS) en América Latina. Aunque buena parte de la literatura y de la trayectoria de investigación se ha concentrado en comprender las dinámicas digitales y su relación con el mundo social –sobre todo, desde las redes sociales y plataformas digitales–(Hampton, 2017), nuestra propuesta no privilegia solo esas temáticas, sino que plantea también un abordaje holístico vinculado a las transformaciones en la generación y disponibilidad de información –pública, sobre todo–y el uso que esta puede tener para el aprovechamiento en las ciencias. Aquí damos fundamentos para intentar resolver, con herramientas digitales, preguntas más amplias sobre las dinámicas tecnoeconómicas y sociales, derivadas de la realidad institucional y estratégica en nuestras regiones.

A partir de ello, el trabajo se va a concentrar en un tipo de abordaje sobre los datos digitales de entre tantos otros posibles: el estudio de la información textual o *text mining* (Hearst, 2003; Grimmer, Roberts y Stewart, 2022) que se genera dentro de los diferentes entramados institucionales de los sistemas de CTIS de América Latina. El supuesto central para esta decisión es que a partir de estas fuentes es posible acceder, en una escala y a una velocidad antes impensadas, a elementos claves para entender cómo el conocimiento se produce, se apropia y circula en los contextos latinoamericanos. Asimismo, representa una novedosa herramienta para la toma de decisiones en el espacio público y para la comprensión de situaciones a nivel regional y local.

A partir de reflexionar sobre las potencialidades y limitaciones de estos métodos y fuentes de información, el capítulo propone también discutir sobre las implicancias de su aplicación para este campo de estudio, las ciencias sociales. Así, moviliza en paralelo una reflexión a la vez necesaria y estratégica sobre las interacciones entre investigación social y los procesos de digitalización transversales a dinámicas económicas, culturales y relacionales, entre otras dimensiones. Lo hace buscando

explicar y exponer aspectos concretos derivados de la documentación disponible de forma pública en internet, considerando los alcances y restricciones inherentes a este tipo de abordajes.

El objetivo de este trabajo se centra en la presentación de los métodos de análisis semántico-estadísticos más recientes, y en proponer formas de explotar su potencial y resolver sus principales limitaciones. Para esto, hemos decidido utilizar a modo de ejemplo diferentes tipologías de temas-problemas a la par de estructuras de datos variados para reflexionar acerca del estudio de problemáticas relacionadas con las realidades latinoamericanas.

En un nivel más específico, en este apartado se busca reflexionar sobre tres dimensiones clave. La primera apunta al rol que proponen nuevas corrientes de pensamiento orientadas a utilizar técnicas de análisis semántico-estadístico para analizar coordenadas espacio-temporales de los discursos, poniendo en relieve no solo el rol de los actores y de los conceptos, sino el de la historia y el de los territorios. La segunda pone foco en la necesidad de profundizar en el desarrollo de investigaciones capaces de hacer uso de la información actualmente disponible, sobre América Latina y el mundo en términos de CTIS. Esto adquiere especial relevancia al contemplar escenarios de producción científica incrementales, del crecimiento de repositorios de documentación digital y de la difusión de nuevas formas de expresión socioculturales y técnicas asociadas al uso masivo de las TIC (tecnologías de la información y la comunicación). Por último, se proponen los "rastros digitales" como vías de acceso al estudio de fenómenos sociales, en el marco de discusiones en torno a la potencial saturación en los métodos investigativos tradicionales y de grandes desafíos metodológicos para las ciencias sociales a partir de la emergencia de enfoques 4.0 derivados de la era digital.

Los potenciales usos de la información textual disponible, el aprovechamiento de fuentes digitales -o digitalizables–y su impacto como insumos en la investigación tienen un rasgo que es de especial pertinencia para América Latina. Las limitaciones de recursos para realizar investigación han sido tradicionalmente un fenómeno que atraviesa las realidades de los diversos sistemas de ciencia y tecnología de la región. En ese marco, la condensación de nueva información disponible en internet, derivada de la dinámica cotidiana a nivel institucional, económicoproductivo, social, académico, entre otras dimensiones, pone de manifiesto la importancia de las nuevas herramientas tecnológicas para impulsar estudios e investigaciones que interroguen y visibilicen nuevas

heterogeneidades y convergencias a nivel regional. Esto es especialmente relevante para el estudio de aspectos tan dinámicos como el estudio de la problemática de la CTIS en esta parte del mundo.

Las capacidades analíticas que se abren a partir del uso de nuevos métodos, técnicas y tecnologías impactan en múltiples dimensiones en el estudio de la ciencia y la tecnología en la región. Estas dinámicas tienen un potencial transformador y habilitan nuevas oportunidades en algunos aspectos clave. Primeramente, se destaca la posibilidad de construir nuevos objetos de estudio, más amplios y complejos, que permitan atender áreas de vacancia en contextos particulares. De la mano de ello, se hacen posibles nuevas formas de colaboración en dinámicas de investigación. La apertura de datos, la posibilidad de replicar los análisis, el acceso a herramientas y *scripts* de disponibilidad pública favorecen la posibilidad de desarrollar agendas de investigación comunes e incrementales. Por último, estas nuevas aproximaciones sostienen la posibilidad de desarrollar conocimiento original sobre la región desde la región. De esta forma, la posibilidad de apropiarse de los nuevos rasgos de este paradigma se vuelve estratégica para pensar las posibilidades de desarrollo de nuevas y necesarias agendas de investigación en la región.

Digitalización, información textual y prácticas de investigación. Algunas reflexiones metodológicas necesarias

El adentramiento en las posibilidades de detección de regularidades significativas en grandes cuerpos de texto es una propuesta novedosa. Avanzar sobre ellas desde en un campo como el estudio de los procesos de aprendizaje e innovación brinda a quienes estamos interesados en estos temas la oportunidad de movilizar una reflexión que es teórica y metodológica, pero, a la vez, es sobre todo práctica. Esto sucede, en parte, porque la generación de datos digitales ocupa un lugar cada vez más importante en el mundo en que vivimos y, en ese marco, ha habilitado gradualmente la posibilidad de estudiar mediante prácticas e interpretaciones específicas resultados determinados derivados de la interacción que hagamos con esa información (Heyes La Bond *et al*., 2017). Es decir, las formas en que documentamos, investigamos y comunicamos nuestros hallazgos no están exentas de ser transformadas por lo digital, y plantean, simultáneamente, un cambio en los tópicos posibles, los métodos y las estrategias de investigación.

Por otra parte, y esto resulta particularmente importante para los objetivos que este capítulo persigue, los archivos digitales no son producto del diseño de ningún investigador (Lazer y Radford, 2017). El trabajo sobre datos textuales en espacios institucionales requiere una teorización clara del contexto de enunciación de estos textos y de sus implicancias para las aspiraciones del proyecto de investigación que los analice. La relación entre los fenómenos lingüísticos y sociales –en este caso, entre los archivos de texto y el fenómeno observado– ha sido estudiada por la historia conceptual, destacando un punto central a nuestros intereses: hay condiciones extralingüísticas que son significativas y que pueden ser capturadas lingüísticamente (Koselleck, 2004: 222-223), en este caso, mediante el análisis semático-estadístico.

Estas técnicas que permiten operar sobre grandes bloques de información textual y derivar a partir de ello una interpretación de la realidad no están exentas de limitaciones. Como bien argumenta Franco Moretti (2013: 218-220), el uso de este tipo de aproximaciones lleva aparejado una serie de implicancias relacionadas con que, cuando uno construye una representación de un corpus específico, lo que tiene entre manos es una abstracción de este, ni más ni menos. Las formas de operar sobre los textos y las maneras de lograr establecer un diálogo con las preocupaciones analíticas y los marcos teóricos son un punto insoslayable. Esto ayuda a entender por qué los modelos de análisis del lenguaje pueden ser útiles, incluso si son limitados. Pero, sobre todo, ponen en evidencia la importancia que tienen las estrategias de validación, es decir, las formas de probar que las categorías utilizadas son confiables y relevantes (Grimmer y Stewart, 2013: 269-71; Grimmer, Roberts y Stewart, 2022), aportando piezas fundamentales que construyen la solidez del proceso investigativo.

Relacionado con el punto anterior, vale tener presente que, al momento de diseñar proyectos y programas de trabajo con estos métodos, es importante la reflexión sobre la dimensión y las cualidades políticas que encierra su aplicación. Las abstracciones a las que hacíamos mención implican, como en cualquier trabajo estadístico, la priorización de ciertas formas de ordenar y sistematizar una concepción teórica del mundo. Estos métodos no están exentos de los ciclos que marcan las tecnologías emergentes, especialmente en relación con diversas expectativas y representaciones de las implicancias de las capacidades tecnológicas presentes y futuras (Peine, Spitters y van Lente, 2013: 1616-1617). Parte de este asunto se explica a partir de los nuevos nexos y configuraciones entre las maneras de pensar problemas de la economía y la sociedad, así

como los modos de intervenir en ellas, resignificando las formas estadísticas que han servido para declamar los deberes del Estado y medir su éxito (Desrosières, 2013).

Así, estos métodos amplían el dominio sobre los datos estadísticos y archivísticos. Por eso, permiten movilizar nuevas narrativas y, al mismo tiempo, construir nuevos objetos de estudio (Ruppert y Scheel, 2019). Recuperar las cualidades políticas de estos rasgos implica pensar cómo la información disponible puede ayudar a reflexionar y accionar empíricamente en la configuración de nuevas realidades y de explicaciones del mundo que nos rodea. En este texto intentaremos dar sucesivos ejemplos de cómo estos métodos pueden ser utilizados para visibilizar tendencias, heterogeneidades y convergencias a nivel latinoamericano; y el potencial que esto encierra para los estudios en ciencia, tecnología, innovación y sociedad (CTIS).

El giro que se manifestó en el gradual uso de estos métodos y técnicas, así como en la interpretación política de sus aplicaciones y resultados, requiere de otra salvedad. Esta se da sobre una dimensión meramente práctica. A partir de las herramientas de análisis semántico-estadístico se habilitan nuevas formas de colaboración entre investigadores. En este campo es normal que tanto las herramientas (*scripts*) o los datos resultados de su aplicación puedan ser fácilmente replicados y copiados. Esto ha abierto posibilidades para facilitar la reproducción de resultados y hacer más transparente el proceso de investigación (Lyon, 2016). Estas prácticas, que fácilmente pueden acomodarse bajo el paraguas de la Ciencia Abierta, interpelan al investigador desde la posibilidad de construir agendas colaborativas y aprovechar distintas instancias de interacción para resolver las vicisitudes inherentes a estos análisis. Sin embargo, esto no constituye un mandato para realizar una apertura total en términos de colaboración y acceso a los resultados (Arza y Fressoli, 2017). Más bien, estas posibilidades deben ser evaluadas detenidamente por quienes deseen embarcarse en iniciativas de este tipo.[6] Una valoración de costos y beneficios es central tanto durante el diseño como en la ejecución de las propuestas de investigación.

Hasta aquí hemos elaborado una serie de puntos. Primero, los procesos de digitalización afectan qué y cómo investigamos. Luego, si decidimos

6 Algunas formas de abordar las asimetrías existentes en los procesos de producción y apropiación del conocimiento pueden ser útiles para meditar sobre este tema. Particularmente, una atención a los rasgos del mercado que puedan permitir a terceros apropiarse de los conocimientos generados (Kreimer y Zuckerfeld, 2014).

trabajar con fuentes y archivos digitales, debemos tener en cuenta que son caóticos y no han sido construidos específicamente para lo que necesitamos. De esto se desprende que una reflexión teórica y crítica sobre los datos es necesaria para poder operar correctamente sobre ellos. Reducir su complejidad de una forma que permita explicar los fenómenos estudiados no es fácil, pero sí encierra mucho potencial de transformación e implica una gran novedad. Finalmente, la posibilidad de transparentar, replicar y colaborar es inherente a este tipo de trabajos. Las maneras en las que los investigadores dialoguen con las posibilidades que habilitan las nuevas técnicas es definitorio para la calidad de los resultados de investigación que logren.

Estos diálogos comienzan, generalmente, a partir de una revisión o reflexión sobre las fuentes a las que se tiene acceso. La convergencia que experimentan las redes y sistemas informáticos ha facilitado el acceso a grandes volúmenes de texto que pueden vincularse analíticamente a los intereses del campo de la CTIS. De hecho, existen muchos más datos accesibles de los que uno imaginaría. Dado que es posible transformar la estructura de observaciones y variables en bases de datos, se abren oportunidades sustanciales para integrar múltiples fuentes en una misma investigación. Esto habilita enfoques capaces de jerarquizar la complejidad y las perspectivas multinivel en los estudios que involucren diversas fuentes textuales.

Este tipo de trabajo constituye un importante fundamento que pone de relieve las aplicaciones de las ciencias de la computación al estudio de la realidad social y económica. El desafío para América Latina radica en poder establecer preguntas propias en este nuevo horizonte emergente y, al mismo tiempo, comprender las características que tiene la información textual en la región. El análisis sistemático y a gran escala de colecciones de texto es ahora posible sin el volumen de financiamiento que sería necesario si no se contara con estas herramientas (Grimmer y Stewart, 2013: 268). Por todo lo expuesto hasta aquí, esta posibilidad se vuelve sumamente desafiante y se presenta como una posibilidad para acumular capacidades propias que jerarquicen un perfil latinoamericano en estos temas. Es decir, sostener investigaciones originales sobre temas críticos y con alcances no abordados que eran, hasta ahora, imposibles.

Tradiciones y perspectivas sobre el análisis semiautomático de textos

El análisis de datos textuales aquí delineado excede el estudio de registros derivados de la interacción en redes sociales, o servicios digitales, como ya se ha dicho. La propuesta apunta a una aproximación más abarcativa, que busca aprovechar el potencial de información disponible, especialmente de acceso libre y gratuito, derivada de la existencia de procesos institucionales y sociales que han dejado huellas y rastros en diferentes soportes escritos. La recolección de esta información textual es realizada de forma pasiva a raíz de la digitalización de la interacción social (Hampton, 2017: 176-179), en la que el texto se convierte en un material analítico con gran potencial, poco utilizado en la investigación de nuestra región. De hecho, la convergencia tecnológica y social sobre plataformas digitales ha movilizado que los archivos digitales crezcan exponencialmente, especialmente desde los 2000 en adelante (Hilbert y López, 2011). Asimismo, este proceso fue potenciado por las mejoras en las posibilidades de digitalización de fuentes analógicas. La utilización de software para el reconocimiento de caracteres (denominado OCR, que discutiremos más adelante) abre, por ejemplo, la posibilidad de construir series históricas más amplias para extender el interés sobre un fenómeno determinado.

En cualquier caso, lo importante es comprender que el insumo con el cual se trabajará no ha sido construido específicamente para fines investigativos. En otras palabras, estos datos no son resultado de una indagación ordenada, coherente y robusta sobre un objeto determinado. Son más bien datos que existen y cuya finalidad es variada (registros, archivos, producción científica, difusión científica, etc.) y que han sido creados con fines no asociados con el proceso de investigación que se aplica a partir del análisis semántico. El valor de este material viene dado, entonces, por cómo se trabaje sobre él. Estas fuentes se oponen a los datos con los que generalmente se trabaja en ciencias sociales, que son creados o recolectados específicamente en función de una hipótesis de investigación, o que están constituidos por registros derivados de instrumentos de comprensión de la realidad de orden transversal, como encuestas o censos. La propuesta del análisis semántico-estadístico propone y se potencia del uso de todo lo anterior de forma combinada. En otras palabras, los métodos digitales a los que hacemos referencia se enriquecen y nutren a partir de complementar registros preexistentes, no necesariamente diseñados para ser un insumo de la investigación,

con métodos y fuentes tradicionales que sí han sido construidas a tales fines (Salganik, 2018: 355-356).

Esta operación de unir fuentes estructuradas y sin estructurar está marcada por dos cuestiones. Primero, por la evolución de los métodos y herramientas utilizadas. Y, segundo, por las tradiciones que han movilizado estas evoluciones. La modularidad de las tecnologías digitales ha permitido que las herramientas puedan frecuentemente aplicarse entre distintas tradiciones analíticas. Esto hace que, muchas veces, sea complicado distinguir cómo un trabajo determinado se inserta en una tradición previa. Esto abre, sin embargo, un espacio importante para la experimentación y la búsqueda de nuevas aplicaciones analíticas. A continuación, haremos una breve reseña de las principales avenidas en las que se inscriben los intentos por estudiar grandes cantidades de información textual.

El análisis de texto y la complejidad de su evolución histórica

A nivel general es importante entender que realizar investigación social con grandes volúmenes de texto es parte de las estrategias analíticas que han surgido a partir del desarrollo de la llamada "ciencia de datos" (Donoho, 2017). La variedad de posibilidades al utilizar los textos como datos se sostiene en una amplia gama de herramientas estadísticas para procesar el lenguaje y relacionarlo con su contexto de enunciación (Grimmer y Stewart, 2013: 268; Grimmer, Roberts y Stewart, 2022). Donoho (2017: 750) ha sintetizado esto al decir que el giro más importante de estas metodologías que se propone es el de dejar de analizar datos y empezar a aprender de ellos. Esta búsqueda ha dado lugar a un escenario que dista de ser homogéneo, en el que se han identificado variaciones notorias en los roles, las instituciones y las trayectorias de los científicos de datos (Fayyad y Hamutcu, 2020). Esto, entre otras cosas, implica que el uso del mismo término no es necesariamente una garantía de bases epistémicas compartidas entre diversos abordajes. Para explicar esto veremos a continuación tres áreas que alimentan el tipo de análisis que aquí proponemos: el análisis de coocurrencia de palabras, el rol de la lingüística computacional en la construcción de corpus de análisis y, por último, el desarrollo de técnicas de procesamiento del lenguaje natural (NLP, por sus siglas en inglés).

A pesar de la heterogeneidad previamente señalada, pueden distinguirse algunas tradiciones de importante significatividad, sobre todo a los fines que persigue este capítulo. Vale mencionar, entonces, las tradiciones de la CTS (ciencia, tecnología y sociedad) y en especial los estudios de "cienciometría cualitativa" como una vertiente de relevancia (Abdo, Cambrosio y Cointet, 2020). Los trabajos seminales que compararon la coocurrencia de palabras en la literatura científica han sido centrales para el análisis de grandes corpus textuales (Bauin *et. al*., 1983; Callon, Law y Rip, 1986). Esto ha sostenido, por ejemplo, la construcción de un tipo de métrica que centra la comparación en la frecuencia neta de las palabras con respecto de las frecuencias esperadas. A partir de asociaciones de este tipo, el estudio de los vínculos semánticos ha sido la base de la adaptación de la teoría del actor red al estudio de los textos científicos (Callon, Courtial y Laville, 1991).

Por otro lado, se encuentra la lingüística computacional. Retomando ideas de Bally, de Saussure y Sechehaye (2003) sobre la organización dicotómica entre lengua y palabra, entre sistema y sintagma, entre sincronía y diacronía, esta línea de trabajo ha buscado entender la distribución y las relaciones paradigmáticas dentro de los textos a partir de usar herramientas informáticas de tratamiento y análisis.

El estudio del lenguaje en grandes volúmenes y en su contexto de enunciación marca el énfasis empírico de esta línea de trabajo. Aunque gran parte del debate de conceptos y sus desarrollos fundamentales hayan tenido lugar a partir de los años cincuenta, las herramientas para construir, catalogar y buscar dentro de grandes corpus de textos se desarrollarían a la par del aumento de las capacidades de cómputo durante los noventas (Hardie y McEnery, 2013). El estudio cuantitativo de la evolución de la "literatura mundial" emprendido por Moretti (2013) y su "lectura distante" puede contarse como un destacado hito dentro de esta área de estudios. A partir de sus mapas, series y árboles de contenidos, Moretti ha inspirado otros intentos por capturar y operacionalizar conceptos amplios y difusos como fuera el de "literatura mundial" antes de su trabajo. El giro a destacar aquí es que conceptos y problemas que por mucho tiempo fueron inabarcables empíricamente, a partir de estas contribuciones, pueden ser capturados mediante la implementación de nuevas técnicas basadas en la informática.

El desarrollo de las herramientas de detección de lenguaje natural (NPL, por sus siglas en inglés) es tributario de algunos de estos avances y a la vez complementario. El énfasis en encontrar, sintetizar y presentar

de forma exitosa y sucinta grandes volúmenes de texto ha sido el gran eje de los estudios de la lingüística computacional (Hirst, 2013). El énfasis, aquí, ha sido el uso de la estadística para detectar y definir términos representativos de un corpus textual. La caracterización del texto se ha vuelto un problema matemático y matematizable. Las técnicas de NPL han sido un gran salto para estos intereses.

En esta misma línea, los estudios desde la teoría de la información han complementado los intereses lingüísticos. Así, comprender las similitudes y diferencias de los textos de un corpus ha implicado un desafío en términos computacionales (Corley, Mihalcea y Strapparava, 2006; Inkpen e Islam, 2008). Las aplicaciones de estas técnicas van mucho más allá de los intereses de este capítulo, pero ejemplifican el tipo de modularidad de las tecnologías digitales y su posibilidad de ser adaptadas a distintos usos. La detección de núcleos temáticos y relaciones entre expresiones lingüísticas constituye la base de múltiples tecnologías emergentes como la inteligencia artificial y el *machine learning*.

Recapitulando, podemos decir que la cienciometría y el estudio de coocurrencia de palabras se han enfocado en comprender las dinámicas de CTIS a partir de describir interacciones, evoluciones y factores destacados de estas. La intersección entre lingüística y computación se ha focalizado en hacer emerger cualidades estructurales en grandes corpus de texto y destaca también la importancia del lenguaje en uso. Por último, la teoría de la información posibilita el desciframiento de la estructura semántica de los textos y muestra su heterogeneidad interna. Estas herramientas han permitido algunos movimientos similares a los de Moretti (2013) en otras áreas de investigación. Haremos una breve reseña de las que creemos que son afines al estudio de las dinámicas de ciencia, tecnología, innovación y sociedad.

Una mirada sobre las tradiciones de investigación digital

En historia, el llamado "giro lingüístico" retomó el interés por entender cómo los seres humanos se relacionan entre sí y construyen su mundo a partir del lenguaje (Delanty e Isin, 2003: 170-172). Las preocupaciones sobre los conceptos han ocupado a grandes referentes de la historia social, como Skinner (1969) o Foucault (2017). La perspectiva de Koselleck (2004) señala direcciones afines a los planteos de este trabajo, ya que pone el eje en comprender cómo algunos conceptos específicos se insertan en

estructuras lingüísticas más amplias, como lo son los corpus textuales de gran tamaño. Las transformaciones sociales y económicas se ven reflejadas en conceptos emergentes, por ende, mientras que los registros escritos reflejan distintas capas de esos procesos, también permiten acceder al estudio de los cambios en los procesos de construcción y socialización de nuevas ideas. Lo anterior se torna especialmente relevante ya que ciertos conceptos tienen roles preponderantes en la organización de las actividades científico-técnicas y en los propios flujos de conocimiento (Rip y Voß, 2013; Bensaude-Vincent, 2014), y el reconocimiento de su dinámica es uno de los ejes centrales para los métodos basados en el análisis semántico-estadístico.

Los tópicos que se desprenden de la concepción de "humanidades digitales" han servido para apuntalar trabajos de interpretación y estudio de la realidad humana en sus múltiples facetas. Esto implicó establecer métodos y procedimientos con un tipo de sistematización muy diferente a la utilizada por enfoques previos. Los avances interdisciplinarios que comenzaron a finales de la década del cuarenta terminaron por conformar una tradición propia (Hockey, 2004) dentro de esta línea de estudios. Los esfuerzos erráticos y poco coordinados se extendieron hasta principio de los años setenta e intentaron resolver problemas de procesamiento masivo, las variaciones en la escritura y la automatización del análisis morfológico o "lematización". La consolidación, que duró hasta mitad de los años ochenta, procuró el desarrollo de espacios para diseminar los hallazgos que eran resistidos por las comunidades tradicionales del campo. Los nuevos desarrollos, entre mediados de los ochenta y principios de los noventa, buscaron herramientas de la lingüística computacional para expandir el trabajo realizado. Pero fue con el desarrollo de las tecnologías de la información y comunicación, desde mediados de los años noventa, que el campo tomó las características actuales. Las nuevas formas de representación y la plasticidad de las formas digitales (Berry, 2012), así como la abundancia de información, y no su escasez, se convirtió en un desafío analítico, metodológico y conceptual.

Los intentos por capturar y entender el dinamismo de los objetos digitales inspiraron un área de trabajo sintetizada por Rogers (2013) bajo el rótulo de "métodos digitales". Lo que ha caracterizado a esta línea de trabajo es el intento de pensar en paralelo los métodos y las transformaciones propias de lo digital. Es decir, diseñar herramientas para observar qué está pasando dentro de plataformas digitales de socialización. Para los objetivos de este capítulo, son sumamente relevantes las herramientas

que resultan de esta tradición, ya que permiten capturar y enriquecer múltiples fuentes de información, sobre todo en plataformas que son nativas de internet y de las esferas digitales que median de forma incremental la interacción humana.

Por último, existe un área de interés definida por los intentos de hacer ciencias sociales a partir de la masiva disponibilidad de datos que dejamos en soporte digital como resultado de actividades cotidianas (Adamic *et al*., 2009). La utilización de "rastros digitales" para la investigación social es una de las características del trabajo de predicción y asociación que hacen grandes corporaciones privadas –como Google, Yahoo o Facebook– y agencias de gobierno. La utilización de estos datos y el aprovechamiento de su minucioso detalle proveen información tanto de la estructura y el contenido de las diferentes interacciones como de su variación geográfica y temporal.

Este espacio de trabajo ha surgido de forma más reciente, a la luz del desarrollo de la *big data* definida a partir de las llamadas "3 V": el volumen de información, la velocidad de procesamiento y la variación de la estructura de los datos. El objetivo, aquí, es sobre todo capturar el potencial que la era digital propone para la investigación social (Salganik, 2018). A diferencia de las humanidades y los métodos digitales, la vocación predictiva de este tipo de trabajos es destacable.

Vale recordar que las principales tecnologías que sostienen estos programas de investigación son modulares y tienden a ser adoptadas en distintas iniciativas investigativas de forma transversal. Esto diferencia este tipo de métodos de tradiciones anteriores, en las que lo teórico y metodológico avanzan a ritmos sincronizados. Por eso, y tomando en cuenta la mencionada modularidad de las técnicas digitales de investigación textual, a continuación, hemos incluido ejemplos de aplicaciones en el campo, una descripción de las herramientas y algunos desafíos propios de esta área emergente.

Ejemplos de aplicaciones en el campo

El estudio de datos desde el análisis semántico hace uso de la estructura textual para caracterizar bloques de información mediante estadísticas descriptivas (Griffiths y Steyvers, 2002; Lieber, 2009; Fahmy y Gomaa, 2013), identificar patrones mediante análisis econométricos (Ghose, Ipeirotis y Sundararajan, 2007; Baker y Fillmore, 2010) y, de forma incipiente,

abordar cuestiones más complejas como redes de actores y de conceptos mediante SNA (Social Network Analysis) e identificación de SW (*small worlds*) (Collins y Quillian, 1969; Strogatz y Watts, 1998; Blum *et al*., 2003), o apreciaciones y toma de posiciones mediante técnicas de *sentiment-analysis* (Collier y Mullen, 2004; Alani *et al*., 2014; Bandyopadhyay *et al*., 2017).

La posibilidad de avanzar en la caracterización de imaginarios sociotécnicos (Vera, 2016) mediante el uso de información documentada, en particular la relativa a los patrones de construcción del conocimiento científico y de documentación clave para el direccionamiento de los sistemas de ciencia y tecnología y las estrategias de desarrollo (territorial, económico y social, entre otras dimensiones), resulta clave como una faceta metodológica del análisis en el marco de sistemas de CTI y de CTS.

Descripción del método y de las herramientas para su aplicación

El análisis semántico ha surgido como una teoría novedosa para el procesamiento y representación de la información y del conocimiento. Este método permite la puesta en relieve y la representación contextual de las palabras mediante cálculos estadísticos aplicados a un determinado objeto textual o corpus de texto (Dumais y Landauer, 1997), permitiendo destacar un tipo de conocimiento de orden global, inducido indirectamente de los datos de coocurrencia locales en este corpus de texto, constituyendo un metaanálisis de orden explícito genérico.

A nivel operativo, las técnicas utilizadas se apoyan en el lenguaje de programación R de código abierto y acceso gratuito, aunque los elementos expuestos en esta sección pueden trasladarse de forma análoga a Python.[7] En particular, se utilizan dos estrategias para la aproximación al análisis de los datos. Por un lado, un tipo de análisis orientado a la construcción y estructuración de la información, que puede implementarse con R y R-Studio, potenciado por librerías como dplyr y stringr, especialmente diseñadas para la manipulación de datos y de datos de texto. Sumado a ello, pueden agregarse paquetes específicos que ofrecen una potente opción del estado del arte. En particular, pueden implementarse análisis

7 Muchas de las librerías y paquetes disponibles para R tienen su correlato en Python, permitiendo implementar funciones similares, con la salvedad de las técnicas de organización y manipulación de datos, que se corresponden con cada lenguaje de programación. También, puede ser necesaria una reflexión sobre las implicancias éticas de posibles usos de los resultados (Salganik, 2018: 281-354).

de *text mining*, así como análisis semántico-estadístico, análisis semántico latente, y de sentimientos, mediante el uso de paquetes de acceso libre como tidytext, tidyverse y glue que, combinados, ofrecen potentes herramientas, gratuitas y de código abierto. La representación puede llevarse adelante con paquetes como qplot, ggplot2, Plotly, igraph y SNA. Por otro lado, se propone también la utilización de la plataforma online CorTexT(www.cortext.com) desarrollada por IFRIS e INRA para la realización de estudios de *open data* orientados al análisis socioeconómico de la ciencia, la tecnología, la innovación y la producción de conocimiento. Las ventajas de esta plataforma se centran en su interfaz que, aunque restringe las posibilidades de personalización, omite procesos de programación que requieren curvas de aprendizaje más pronunciadas por parte del usuario.

Algunos desafíos para el análisis semiautomático de textos

A modo de cierre de este rápido repaso por las trayectorias técnicas y las tradiciones analíticas que marcan los tipos de abordajes sobre el texto como dato, es importante subrayar los desafíos que investigaciones de este tipo encierran en la actualidad. En primer lugar, están las diferencias entre lo que han sido las "dos culturas" centrales en estos desarrollos: las ciencias de la computación y las ciencias sociales. DiMaggio (2015) explica que esta diferenciación se ha reducido notoriamente como resultado de la estrecha colaboración a la que nos hemos referido en los dos apartados anteriores. Sin embargo, algunas particularidades distinguen los objetivos y los usos de las herramientas de análisis computacional de textos. Resulta importante no perder de vista estas diferencias para reflexionar sobre ellas al momento de diseñar los flujos de trabajo y las estrategias de investigación.

DiMaggio plantea algunos argumentos que, a primera vista, parecen contraintuitivos. Primero, los investigadores sociales usan, generalmente, modelos no supervisados[8] para explorar los corpus. En contraste, los informáticos prefieren tener formas de validar y enriquecer sus resultados haciendo explícitos los modos de validación utilizados. Las formas de validación constituyen un punto clave que retomaremos en breve.

8 Ejemplos de estos métodos son aquellos que no parten de ningún tipo de clasificación previa y se centran en "hacer emerger" estructuras subyacentes en los textos analizados.

En segundo lugar, los investigadores sociales tienden a explotar intensamente los algoritmos y herramientas existentes, mientras que el énfasis de los informáticos está en el desarrollo de nuevos modelos. Los programadores pueden desarrollar soluciones más rápido de lo que los investigadores sociales pueden aprenderlas y adaptarlas a sus necesidades. Las curvas de aprendizaje importan y, por ende, explican muchas de las trayectorias en la utilización de estos métodos.

Por último, y retomando el primer punto, los informáticos parecen confiar mucho más en el juicio de los seres humanos que los sociólogos, economistas y analistas sociales. La preocupación por los sesgos que la participación humana puede tener es un aspecto que preocupa en la investigación social y vuelve atractiva la validación estadística. Justamente, el desafío para utilizar métodos de análisis textual para estudiar problemáticas específicas, como las CTIS, radica en encontrar formas para evaluar y explicar claramente la elección de un modelo en torno a un corpus textual y, sobre todo, entender para qué puede servir el aporte complementario de fuentes cualitativas en la comprensión y abordaje de un problema en particular.

Esto implica, necesariamente, reconocer los límites de estos métodos (Baya Laffite *et al*., 2014). El volumen de información no hace necesariamente a la calidad de esta, muchas veces el valor está en la variedad de estas fuentes y en su representatividad. Luego es clave reflexionar sobre las prácticas que han dado luz a los datos analizados y admitir que estos rastros digitales son artefactos sujetos a las mismas restricciones que cualquier otra construcción humana. Esto, a la larga, potencia la calidad de los hallazgos, no la disminuye. Finalmente, debe reconocerse que lo digital no es automático. Por eso, la intervención del investigador es crucial y de aquí viene el concepto de semiautomático que hemos utilizado a lo largo de este capítulo. Los esfuerzos para limpiar, transformar e interpretar los datos textuales agregan valor a los resultados y los hacen más valiosos.

Entonces, el desafío de la validación de los resultados de análisis semiautomático de texto es lo que determina la calidad de los resultados analíticos. El énfasis en la complementariedad entre momentos de exploración y modelización cuantitativa, y momentos de validación y enriquecimiento cualitativo, debe ser siempre tenido en cuenta. El análisis cuantitativo permite identificar regularidades, capas ocultas en los textos, que pueden ser excelentes puntos de entrada para otro tipo de análisis más detallado que los ponga en valor.

Definiciones del método y estrategias de implementación

La aproximación hacia los datos y su estructura: cuestiones clave a tener en cuenta

En términos generales, la primera instancia del análisis semántico consiste en la definición de fuentes de información y tipos de datos sobre los cuales realizar el análisis. Es importante destacar que, en relación con lo anterior, la definición del objeto de estudio, la pregunta de investigación, hipótesis y contextualización del tema-problema serán un requerimiento clave, tanto para el desarrollo de estas metodologías, como para cualquier abordaje empírico en ciencias sociales. Respecto a las técnicas de análisis semántico, se deberá tener certeza del tipo de dato sobre el cual se desea indagar, prestando atención no solo a sus características agregadas (entre otras, el tipo de documento, idioma, estilo de prosa, tipología de oraciones o sentencias, palabras clave, estructuración del texto en general y de sus metadatos, etc.), sino también sobre cómo este dato se organizará dentro de una base de datos o *dataset*.

La proyección del corpus textual en un *dataset* puede resultar un ejercicio complejo y, muchas veces, requerirá de pruebas de ensayo y error en una escala piloto para garantizar que el resultado sea el deseado. En este sentido, las formas de articular los datos pueden agregarse en tres grandes grupos: una base de datos integral, con todas las variables que buscan analizarse en la investigación (por ejemplo, los campos de un *curriculum vitae*); un conjunto de bases de datos más pequeñas articuladas entre sí por vectores conectores (por ejemplo, segmentar diferentes estructuras de datos para diferentes tipologías de documentos que definen un mismo objeto de estudio); o un número de vectores, articulados por identificadores clave capaces de conformar subsets de información según la necesidad.

La elección del tipo de estrategia para organizar y utilizar los datos dependerá, a su vez, de las capacidades de procesamiento y de las herramientas técnicas que se utilicen en la investigación. En este sentido, el uso de una computadora personal puede resultar cómodo y versátil para estudios de baja escala, pero sin dudas presentará limitaciones en trabajos que busquen abordar datos extensos en volumen o en heterogeneidad. Solo para esbozar brevemente el tipo de problema que esto representa y su relación con la estructuración de datos, vale decir que en la gran mayoría del software de procesamiento de datos el almacenamiento de

la información tiene estructura matricial, por lo que al incrementarse el tamaño de la matriz (especialmente a nivel horizontal, es decir, en la cantidad de variables existentes) suelen encontrarse límites tanto a nivel del software (por ejemplo, Stata/SE tiene un límite de cinco mil variables) como a nivel del hardware (puntualmente relacionados con la disponibilidad de memoria RAM, generando procesos de *swapping* -escritura en el disco en lugar de la memoria–que ralentizan sensiblemente el procesamiento). Las soluciones de *cloud computing* o de clústeres de procesamiento suelen ser una alternativa capaz de sortear las limitaciones anteriores, aunque también conllevan costos en su uso (ya que generalmente se rentan por hora a proveedores de escala global). La utilización de matrices completas o vectores para la estructuración de datos dependerá, en suma, de lo siguiente:

1. Qué tipo de datos se busquen analizar.
2. El software y la técnica que se utilicen.
3. El tipo de hardware del que se disponga.

De llevar adelante análisis de gran cantidad de datos en una computadora personal, será prioritario considerar el requerimiento de trabajo con vectores en lugar de matrices en la estructuración de datos.

Por último, las observaciones anteriores son válidas para los abordajes más difundidos de análisis textual desde una perspectiva cuantitativa: *data scraping*, *data mining*, *text mining*, *natural language processing* y análisis semántico, solo por mencionar las aproximaciones que, actualmente, son de uso más extendido.

Las diferencias entre estos enfoques consisten en el tipo de información analizada, el procesamiento al que se someten los datos y, especialmente, en la indagación que se tiene por objeto de estudio. A modo agregado, la mayor distinción entre estas técnicas puede identificarse entre la familia del *data mining* y la del análisis semántico: en el primer caso, el objeto de estudio indaga sobre el texto explícito, aplicando técnicas de descripción estadística, de análisis cuantitativo y de modelización; en el segundo, se busca estudiar categorías latentes dentro de los datos, por lo que los análisis requieren ciclos de aprendizaje y ajuste de las categorías para mejorar las predicciones, requiriendo técnicas de *machine learning* y de procesos iterativos de mejora de los algoritmos encargados del procesamiento de los datos. En los siguientes ejemplos, nos centraremos

especialmente en la primera categoría. Hacia el final del documento compartimos algunas observaciones sobre el segundo caso.

Descripción del flujo de trabajo

El flujo de trabajo para la implementación de métodos de análisis semántico sobre grandes cantidades de texto suele dividirse en cinco grandes etapas:

1. *Web crawling* y *data wrapping*
2. *Data parsing* y la selección de variables clave
3. Estructuración de datos
4. Procesamiento inicial
5. Procesamiento específico

Como se aclaró anteriormente, todas estas instancias responden a una hipótesis de trabajo claramente definida, que determinará las estrategias a aplicar en cada uno de los bloques de trabajo.

El *web crawling* y el *data wrapping* son procesos de recolección de datos desde fuentes disponibles en internet. El primero consiste en explorar la existencia de información específica, en general, en búsquedas abiertas asistidas por buscadores (como DuckDuckGo o Google). El segundo, es una aproximación que se aplica cuando los datos sobre los que se desea realizar el análisis no están sistematizados en forma de base de datos (es decir, ordenados por tipos de información y de variables, construidos con la finalidad de utilizarlos para realizar análisis cuantitativo). Así, el *data wrapping* se centra en la adquisición de información disponible online, que puede pertenecer a diversas fuentes o presentar estructuras heterogéneas a pesar de estructurarse en una fuente de información unificada. En algunos casos, técnicas de automatización de la recolección, mejor conocidas como *web scraping*, pueden ser requeridas en este punto (Marres y Weltevrede, 2013). En otros, la disponibilidad de interfaces de aplicación o API puede facilitar las consultas a los sitios web para acceder a datos determinados. Los cambios suscitados a partir del uso opaco de estos datos han restringido el uso de API y han hecho que las técnicas e

instrumentos de *scraping* tomen nuevamente una creciente importancia (Digital Methods Initiative, 2020).

En este sentido, tanto los datos explícitamente visibles en un sitio web o documento como sus metadatos, es decir, la información almacenada en el documento que suele caracterizar y organizar diversos aspectos de este, pueden ser capturados con técnicas de *wrapping*. Los metadatos tienen la ventaja de ofrecer información estructurada acerca del documento –que se aloja en el propio código fuente de este– y que es útil para su interpretación cuando son abiertos por un software específico. Un sitio web, un documento PDF o un documento Word, disponen de metadatos en su código. Entre otras ventajas, haciendo uso de esta capa de información se puede acceder a detalles del autor, la fecha de creación, el título del documento, etc., aprovechando una estructura de código sistemática y organizada, evitando así la necesidad de realizar un análisis y una categorización exhaustiva del documento en sí para obtener una descripción inicial.

Las técnicas de *wrapping* suelen combinarse con algoritmos de diversa complejidad que automatizan el proceso de consulta. Esto dependerá de las características de la fuente de información (por ejemplo, la estructura y lenguaje de programación con el que fue construido un sitio web), del tipo de documento u objeto de estudio, y de la naturaleza de la información sobre la que se pretenda realizar el análisis. Entre las posibilidades más usuales y difundidas se destacan *scripts* en formato *batch* o *bash*, según la plataforma de trabajo –Linux, Mac o Windows–, o aproximaciones más complejas que hacen uso de lenguajes como Ruby o Python mediados por sintaxis Regex (Bandaru, Moyer y Radhakrishna, 2012). Para hacer frente a esto existen algunas soluciones capaces de ofrecer interfaces de usuario que omiten la manipulación de código, como extensiones orientadas a la automatización de acciones en navegadores web (aunque, si las acciones que se buscan realizar son poco regulares, la necesidad de recurrir a herramientas de programación se torna importante).

El resultado del *data wrapping* es un set de datos estructurados. Podríamos sintetizarlo como documentación organizada, que tiene la ventaja de estar disponible offline para su almacenamiento y manipulación. Usualmente, se trata de archivos derivados del análisis de sitios web (HTML para mayor facilidad de procesamiento) o de documentos de texto accesibles de forma pública. Dado que en general estos datos no disponen de ningún tratamiento, tanto la información que contienen como su formato suelen no ser óptimos para un análisis cuantitativo.

La siguiente etapa de trabajo se basa en la organización y selección de información mediante técnicas de *data parsing*. Este proceso consiste en el recorrido algorítmico de los documentos sobre los que está orientado el análisis, con la finalidad de eliminar información excedente o seleccionar algunos componentes particulares dentro de la estructura de datos (Barbu *et al*., 2009). Usualmente, este conjunto de técnicas se aplica sobre documentos disponibles offline, dada la comodidad para su manejo y la velocidad que implica trabajar sobre archivos locales.

Existen muchas estrategias para implementar la limpieza de datos. El uso de *scripts* es el método más difundido, efectivo y versátil. Los lenguajes a los que se recurre se basan en técnicas de *bash* (lenguaje de consola de sistemas Unix, disponible para Linux o Mac nativamente, e instalable en Windows mediante fuentes gratuitas y abiertas), en particular aplicando programas como grep, sed, awk y fórmulas Regex. Esta aproximación puede realizarse también dentro del entorno de programación R (o dentro de R-Studio, independientemente de la plataforma sobre la que esté funcionando el software) a partir del comando "system()". Las limitaciones del uso de *bash* están determinadas por la complejidad del documento y por la afinidad de cada investigador o investigadora con estas técnicas.

La finalidad de aplicar *data parsing* involucra un doble objetivo: por un lado, el de seleccionar los datos precisos con los que se desea trabajar y, por otro, el del ordenamiento de estos en una matriz o *dataset*, en vectores simples o en bloques de datos (que usualmente son matrices de dos columnas) para luego ser usados en los análisis preliminares de orden descriptivo. Por supuesto, esto último estará estrechamente relacionado con el tipo de dato con el que se trabaje y con el proceso de selección que se haya llevado adelante. Los procesos de *parsing* suelen ser iterativos, es decir que su avance consiste en diferentes pasos acumulativos de un orden de complejidad similar. Esto involucra la ventaja de aplicar líneas de código muy simples de forma sucesiva cambiando levemente su estructura, en lugar de plantear el proceso en un solo paso de código (Berant y Herzig, 2019). Suelen encontrarse ejemplos de esto al utilizar los metadatos disponibles en fuentes HTML o XML, en las que la estructura del código fuente suele ofrecer facilidades para la selección de una porción de texto: de estar interesados en el título de una sección, es usual que cadenas de texto bajo la etiqueta "<Title>" (o denominaciones similares) se encuentren en la misma línea o en la anterior al título buscado. Esto permite realizar el mismo proceso de selección de datos consecutivamente, en un paso seleccionando la etiqueta identificatoria

y en el otro, el texto en sí. Como resultado de esto, obtendremos una serie de datos depurados en variables clave, que responderán la pregunta de investigación en análisis.

La estructuración de los datos, como se comentó anteriormente, es fundamental para tener en cuenta a medida que la cantidad de información con la que se construye el *dataset* se incrementa. En particular, dentro de las técnicas de análisis semántico pueden diferenciarse dos tipologías de información: las variables de caracterización, que suelen encontrarse en los metadatos de los datos analizados, y las variables del corpus textual, que contendrán el objeto de análisis central de la investigación. Dependiendo del objeto de estudio o de las fuentes utilizadas, el corpus puede variar enormemente en su complejidad y tamaño. Por ejemplo, en aproximaciones de análisis bibliométrico el corpus estudiado suelen ser los títulos o *abstracts* de artículos científicos. En estudios de análisis semántico más complejos, la expansión de los datos puede basarse en secciones seleccionadas (desde la introducción a las conclusiones) o incluso involucrar el texto completo (por ejemplo, si el corpus tiene que ver con un análisis de normas o de actas).

A mayor tamaño y complejidad del corpus, mayores serán las dificultades de construir un único *dataset* en el que convivan variables estructurales, derivadas de los metadatos, y variables relacionadas con el texto a analizar. Desde las aproximaciones tradicionales para el análisis cuantitativo, la disponibilidad de un *dataset* integrado es un requerimiento deseable. Sin embargo, el uso de R (y R-Studio) o de Python ofrece soluciones que otros paquetes estadísticos no (especialmente si pensamos en las características de SPSS y Stata). Las ventajas que estas aproximaciones ofrecen al trabajar con "data-frames" relajan muchas de las restricciones más sensibles al momento de operar con grandes cantidades de datos. Estas ventajas se centran en la capacidad de alojar porciones de datos a nivel vectorial para su análisis, sin necesidad de recurrir a la carga en memoria de toda la matriz de información.

En los análisis más complejos, la estructuración de datos y la selección de variables clave responden a dos lógicas, primeramente, técnicas de refinamiento del texto y, seguido de ello, técnicas de análisis basadas en hipótesis conceptuales. En general, el primer procesamiento de datos suele orientarse a estadísticas descriptivas del corpus analizado, mientras que el segundo procesamiento aplica técnicas puntuales basadas en hipótesis de trabajo y en el tipo de análisis conceptual abordado.

Así, las técnicas de primer procesamiento pueden ser consideradas de carácter transversal. Como sucede con otras aproximaciones de análisis cuantitativo, en esta etapa es crucial la presentación clara de la estructura de datos, unidades de análisis y dimensiones clave que caracterizan las relaciones que se elaboren en un análisis específicamente vinculado a la hipótesis de trabajo. Entre ellas, se destacan el tipo de documentos analizados, considerando la heterogeneidad de clases, formatos y estructuras. Esto implicará la posibilidad de un análisis del corpus textual que permita caracterizar aspectos de orden estructural, generando mayor precisión en la interpretación de resultados. En lo que refiere al análisis del corpus, nos referimos a métricas básicas respecto de tipologías, densidades y frecuencias de términos clave y de términos omitidos.[9] Así, esta etapa se sintetiza en la presentación de categorías como *tokens* y *lemmas* (palabras seleccionadas para el análisis, homogeneizadas –mayúsculas, acentos, caracteres especiales, etc.–y para el caso de los *lemmas*, llevadas a su raíz textual;[10] estructuradas como observaciones singulares dentro del set de datos) en tablas de frecuencias, análisis de redes, gráficos u otras técnicas que faciliten la exposición de la información resultante (Foltz, Kintsch y Landauer, 1998).

La etapa siguiente consiste en el procesamiento específico del corpus textual, en el que el análisis se realizará de forma directamente vinculada a la(s) hipótesis de trabajo (Foltz *et al.*, 1998). Usualmente, estos procesos involucran técnicas de mayor complejidad, como enfoques de coocurrencia (Hofman, 2013) y otras técnicas para medir la asociación semántica (Farahat *et al.*, 2005), como el IRR (Iterative Residual Scaling) (Ando y Lee, 2001), el LPI (Locality Preserving Indexing) (Cai, Han y He, 2005; Danilevsky *et al.*, 2012) o modelos de espacios terminológicos también conocidos como Word Space Models (Peirsman, 2008; Geeraerts *et al.*, 2015), entre otros. Entre estas técnicas, las aproximaciones denominadas *análisis de sentimientos* (Liu, Wang y Zhang, 2018) combinadas con estrategias de *machine learning* (Alsadoon *et al.*, 2019) han proliferado en los últimos años.

9 Los términos omitidos son un elemento crucial del análisis. Si bien existe una multiplicidad de librerías disponibles para realizar la remoción de conectores gramaticales o *stopwords*, según el lenguaje del texto analizado (librería *nltk* en Python o librería *stopwords* en R, entre otras), en todos los paquetes para uso científico está disponible la opción de eliminar términos de poco interés, incluso si no pertenecen a la lista de conectores gramaticales.
10 Un ejemplo de esta técnica implica la eliminación de tiempos verbales, transformando todos los casos al infinitivo, habilitando así una cuantificación más clara y precisa.

En general, los modelos de procesamiento específico apuntan a evidenciar relaciones complejas dentro de una estructura textual dada, abordando cuantitativa e iterativamente diferentes niveles de esta. Por ejemplo, es frecuente realizar análisis que luego de caracterizar los términos más frecuentes (procesamiento inicial), realizan una cuantificación de las palabras más cercanas a estos, seguida de técnicas de categorización basadas en aprendizaje iterativo o *deep learning*, que ofrece como resultado la capacidad de generar análisis de superficies (Araque *et al*., 2017) y análisis de sentimientos (Ain *et al*., 2017).

A partir de estas técnicas pueden explorarse relaciones complejas entre términos clave dentro de un corpus textual. La capacidad iterativa de muchos de los algoritmos que median este tipo de abordajes (por ejemplo, en paquetes como NLP o Syuzhet en R) en combinación con la disponibilidad abierta de su código fuente, hacen de estas aproximaciones potentes herramientas capaces de ser adaptadas a necesidades específicas de cada investigación.

Al respecto, es importante destacar una dimensión crítica del análisis semántico computacional. Todas las técnicas enunciadas hasta el momento apuntan a implementar, de forma más o menos automatizada, diferentes caminos para la estandarización, categorización y tratamiento de grandes cantidades de texto. Por ejemplo, pueden servir para interpretar la coherencia de los textos (Foltz, Kintsch y Landauer, 1998), sintetizar respuestas abiertas (Alfonseca *et al*., 2005) y para el desarrollo de investigaciones en el ámbito de la comprensión de textos (Kintsch, 2002). En este marco, toda técnica de análisis semántico latente debe ser interpelada desde la dimensión conceptual. Los análisis automatizados pueden ser facilitadores de estos procesos, aunque no necesariamente arrojan los resultados deseados en las primeras instancias. Es por ello que la documentación y revisión sistemática y detallada de cada etapa del procesamiento de datos es crucial para lograr un resultado consistente y validado. En este sentido, las estrategias de validación cualitativa suelen ser un camino no solo deseable, sino también necesario, para abordajes de análisis semántico latente de grandes corpus textuales. Una buena práctica recomendada es la documentación de los hallazgos preliminares en un *log* o registro anotado de manera tal que las operaciones realizadas sean fácilmente rastreables. A partir de esto es más fácil implementar bifurcaciones o modificaciones a las estrategias empleadas para su mejora o reproducción.

Por último, vale la pena subrayar un aspecto central respecto a este tipo de estudios, relacionado con la capacidad de procesamiento de los datos, su almacenamiento y las estrategias para su obtención. La cantidad de material disponible para la realización de estudios de esta naturaleza, como se señala en el próximo apartado, es abundante y heterogénea. Esto puede implicar grandes desafíos para la gestión de los datos disponibles. En este marco, es importante señalar dos dimensiones: la primera, que existen servicios alojados íntegramente en Internet como Watson de IBM, Sentiment Analysis API de MeaningCloud, o Cloud Natural Language de Google, entre otros, que habilitan el procesamiento de grandes cantidades de datos de forma descentralizada (aunque en la mayoría de los casos no son servicios gratuitos); por otro, que las posibilidades de procesamiento en una computadora personal pueden expandirse si el ciclo de análisis está claramente planificado, mediante el tratamiento de segmentos de datos de forma secuencial. Esto representa una técnica relativamente avanzada en el análisis semántico, pero también implica una estrategia eficiente para abordar análisis de cantidades importantes de información.[11] La existencia de plataformas abiertas para el análisis de texto, como CorText, permite realizar algunas de estas operaciones de forma remota. Sin embargo, quien investiga se ve limitado en las posibilidades de adaptar las herramientas a sus necesidades y a implementar aproximaciones específicas.

Aplicaciones en estudios de ciencia, tecnología, innovación y sociedad

En esta sección se presentarán algunos ejemplos de aplicaciones de análisis semántico para estudios CTI en Latinoamérica. Se expondrán cuatro casos, que exhiben diferentes estructuras, alcances y desafíos desde el punto de vista técnico. Más allá de esa heterogeneidad, lo que busca señalarse a partir de su presentación es la magnitud de áreas de vacancia existentes y la potencialidad del uso de técnicas de análisis semántico para su aprovechamiento en estudios de CTIS. Las fuentes de los cuatro casos expuestos son de acceso público y gratuito.

11 Este tipo de abordajes consiste en procesar partes del corpus en lugar de la totalidad. En general este procedimiento no tiene sentido práctico, excepto cuando la cantidad de datos supera la cantidad de memoria RAM del ordenador en el que se estén realizando los cálculos.

La propuesta del capítulo se orienta a presentar de forma analítica diferentes casos en los cuales puede aplicarse el análisis semántico-estadístico para estudios de CTIS. Cada caso presenta particularidades técnicas y de los datos, pudiendo permitir una discusión acerca de alcances y limitaciones de la aproximación sobre la base de casos de investigación concretos. Lo anterior implica que estos ejemplos proveen escenarios particulares, no solo en relación con los temas-problemas sobre los que se circunscriben, sino también en lo que refiere a la naturaleza de los datos, a los límites o alcances de estos y a las técnicas disponibles para el tratamiento y análisis de la información.

El primer ejemplo que se presenta consiste en el análisis de un corpus textual definido, circunscripto, homogéneo en gran medida en su estructura, y de tamaño moderado. Se trata del estudio de reglamentaciones de nivel internacional derivadas de reuniones diplomáticas (actas, tratados e informes de la Comisión Antártica). En particular, se presenta un caso ejemplificativo centrado en la posición de la Argentina y Sudamérica respecto a las relaciones con países y actores (empresas multinacionales, ONG, entre otros) involucrados en la discusión sobre la soberanía antártica. Cada encuentro es registrado en actas que son de acceso público en formato digital. En función de estos documentos de trabajo y sus anexos, es posible construir una base de datos intertemporal desde 1991 hasta la fecha.

El archivo de textos está compuesto por documentos con diferentes formatos que resumen las actas de discusiones e intercambios diplomáticos. Los registros más viejos están en formato PDF, derivados de la digitalización de actas elaboradas de forma analógica, mientras que los más recientes están construidos en el mismo formato, aunque elaborados con un software editor de textos (en general, Microsoft Word). Esto plantea la necesidad de procesar los registros más antiguos para elaborar la catalogación y tokenización de la información en su interior. Para ello, se debe recurrir a un software de detección y conversión de imagen a texto (Optical Character Recognition (OCR)). La calidad de la fuente, especialmente en lo referente a la posibilidad de identificar claramente los caracteres es clave para el éxito del proceso de conversión con software OCR. Existen herramientas avanzadas, apoyadas en servicios cloud e inteligencia artificial (Watson, AI OCR de Google, Azure, entre otras) que permiten implementar aproximaciones de mayor complejidad mediante el Intelligent Data Processing (IDP), para casos en los que la cantidad de datos a procesar sea importante, o se requieran resultados

específicos. Es usual, en la implementación de la conversión OCR —más allá del método utilizado—, la necesidad de realizar revisiones y ajustes manuales para obtener una versión final que satisfaga los requerimientos del análisis en proceso.

Una característica interesante a destacar de este set de datos tiene que ver con su alcance muestral. En efecto, este archivo puede consistir en el universo de discusiones o puede ser segmentado en una muestra (por ejemplo, un período de tiempo específico). La posibilidad de trabajar con archivos que representen el universo en su totalidad no es usual en algunos casos, como se expondrá en los otros ejemplos que se presentan en este apartado. Esta cualidad de los datos permitirá abordar análisis y conclusiones mucho más claras. De ser posible, es deseable trabajar con este tipo de estrategias muestrales.

El procesamiento del archivo se apoya fuertemente en el objetivo general de la investigación. Por ejemplo, puede plantearse el abordaje de ciertas transformaciones intertemporales en los tópicos tratados en las reuniones, para lo que será necesario establecer una unidad de análisis basada en cada acta. De forma análoga puede plantearse algo similar, pero con el foco en las intervenciones de algún país. Entre un objetivo y otro, el cambio central tiene que ver con las capas del análisis. A mayor complejidad y combinación de dimensiones, mayor será el trabajo de catalogación del archivo de texto.

Una cuestión importante a remarcar es la relacionada con las capas del análisis; ya que a medida que el objetivo de la investigación apunta a resultados más ambiciosos, la cantidad de niveles analíticos se multiplica, así como las etapas del procesamiento. Entre las capas básicas del análisis, siguiendo el caso anteriormente presentado, pueden señalarse cuatro dimensiones iniciales que estructuran el procesamiento y el recorte: documento, país, orador, tópico. Por supuesto, lo anterior puede complejizarse y combinarse, considerando por ejemplo el destinatario o destinatarios de cada intervención, la construcción de redes y conexiones entre tópicos y la implementación de análisis de sentimientos (si la intervención está a favor o en contra en una determinada discusión), entre otras técnicas.

En este sentido, el recorte de la investigación resulta una dimensión fundamental. Las técnicas de análisis semántico pueden presentar oportunidades de escalar operaciones, habilitando o facilitando la implementación de diversos análisis de forma progresiva. En otras palabras, y tomando como ejemplo lo anterior, los pasos requeridos para realizar

un estudio sobre redes de tópicos implicarán esfuerzos que no solo contribuirán a dicho fin, y que podrán ser insumo para otras técnicas. Esto puede llevar a un procesamiento de datos difuso, abarcativo y poco preciso, características que se sugiere no estimular en este tipo de análisis. Se sugiere, por tanto, que la definición de niveles, el recorte de la problemática y el procesamiento de datos se estructuren de forma funcional a la hipótesis de trabajo y se acoten a su tratamiento, para evitar el abordaje y tratamiento de datos que luego no necesariamente será aprovechado.

El segundo ejemplo se propone el abordaje de grandes cantidades de documentación relacionada sobre un tópico o tema-problema, con la particularidad de que el universo de documentos se encuentra concentrado en un sitio web (como típicamente sucede con los *journals* y con los sistemas de indexación como Scopus o Latindex). Esto ofrece un recorte específico de los datos, destacando particularmente las contribuciones de una institución, una revista o un espacio virtual, en las que se asume la existencia de un criterio de selección temática, disciplinar y de calidad que atraviesa todas las piezas de texto a analizar. Así, siguiendo esta idea, puede señalarse como caso de estudio el repositorio de la CEPAL, que actualmente cuenta con algo más de cuarenta mil artículos. Mediante técnicas de análisis semántico, pueden analizarse dichas contribuciones, indagando por ejemplo acerca de las transformaciones de los tópicos predominantes a lo largo del tiempo, de los referentes más importantes y de las vinculaciones entre autores, entre otras dimensiones.

El problema central en el que se circunscribe este tipo de estudios versa sobre algunas cuestiones señaladas en apartados previos. En primer lugar, se destaca la cuestión del tamaño de los datos en análisis, ya que de contemplar un análisis integral del repositorio se trata de un corpus textual de alrededor de 100 GB (a 300 KB por artículo). Esto puede plantear importantes problemas en el tratamiento, en especial si se procede bajo una perspectiva de análisis que busque resultados analizando el total de los datos disponibles. En este sentido, y a modo de ejemplo, la identificación términos clave sobre el corpus de 100 GB no es una práctica deseable; en lugar de ello, deben realizarse análisis parciales sobre cantidades de información de menor densidad, y luego sistematizar esos resultados parciales en un nuevo *dataset*. Esto permite sortear limitaciones de hardware y software, incrementa notablemente la velocidad de procesamiento y habilita la posibilidad de realizar el trabajo desde una computadora personal, sin recurrir a servicios *cloud* pagos.

Las ventajas de la aplicación de técnicas de análisis semántico a este tipo de datos son varias. En primer lugar, permite una circunscripción definida del objeto de estudio, limitando claramente la fuente de los datos, su homogeneidad, su pertinencia, y otros criterios que puedan ser relevantes según el caso específico. Adicionalmente, se destacan ventajas respecto de la estandarización de cada documento, en términos de su estructura y formato, facilitando la identificación de secciones y otros elementos clave dentro del texto. Tercero, de elaborarse un análisis parcial del archivo, es posible determinar con precisión qué recorte se seleccionó para trabajar y qué representatividad tiene respecto del universo.

Un tercer ejemplo que es útil desarrollar apunta al abordaje de problemas con fuentes diversas, en las que es difícil o poco factible determinar la extensión del universo y de la muestra que se estudiará. En este sentido, los casos que se enmarcan dentro de este escenario son abundantes y, solo a modo de ejemplo, se señala el estudio de las articulaciones universidad-empresa en América Latina. A diferencia del ejemplo anterior, esta propuesta se centra en la exploración semántica de un tópico de investigación basado en una muestra derivada de un universo difuso. Este ejemplo puede pensarse en línea con los aportes recientes de Brixner *et al.* (2021), aunque en este caso las preguntas de investigación se enfocarán en las características del corpus de literatura latinoamericana (y no de las contribuciones a nivel global que realizan en el estudio citado).

En los casos en los que la demarcación del universo de análisis es difusa, la confección de una muestra confiable resulta clave. Así, muchas de las recomendaciones de este ejemplo apuntan a resolver problemas relativos con la especificación muestral y los criterios para su confección. Este problema resulta ser uno de los más usuales en estudios semánticos de corte exploratorio, o de corte analítico alrededor de un tema-problema específico.

En los últimos años, la literatura ha avanzado en sintetizar las múltiples contribuciones que se realizan alrededor de un tópico a nivel global. Esto involucra el análisis de una cantidad abundante de información que no necesariamente comparte patrones respecto a la estructura de los documentos, sus secciones y su estilo. Dicha heterogeneidad sitúa las iniciativas de esta naturaleza dentro de las más complejas de abordar mediante el análisis semántico.

La fase inicial del tratamiento de este tipo de abordajes debe contemplar que el problema del recorte de datos resulta crucial para una implementación exitosa de la técnica de análisis semántico por dos motivos

centrales. En primer lugar, el universo de contribuciones sobre un tópico de investigación amplio suele ser difuso en sus límites (las múltiples fuentes, espacios de publicación y técnicas de búsqueda son algunos de los factores que pueden influir sobre esto) y, por lo tanto, es extremadamente complejo realizar una circunscripción exhaustiva de las contribuciones.

Lo anterior limita la estrategia de construcción del corpus a analizar y descarta la posibilidad de trabajar de forma explícita con el universo, planteando necesariamente la selección de una muestra como criterio de trabajo. En segundo lugar, y relacionado con lo ya dicho, los criterios de selección de la muestra deben tender a minimizar los problemas anteriormente destacados. Esto es, dicha estrategia debe contemplar factores que sean capaces de circunscribir el problema a un estadio lo más cercano posible a un abordaje exhaustivo, limitando el grado de amplitud en la inclusión de nuevos elementos. En tercer lugar, el rol que ocupan las técnicas y los motores de búsqueda para la identificación de artículos en Internet es determinante. Al respecto, vale la pena señalar que se debe tener presente el uso de múltiples motores, ya que los resultados entre uno y otro varían (actualmente, no es recomendable la omisión de Google Scholar, Semantic Scholar, DuckDuckGo o Microsoft Academic). Finalmente, para el caso del abordaje de problemas relativos a la revisión conceptual, uno de los recortes posibles es el de las contribuciones desde una región en particular, o sobre una región en particular. En nuestro caso, se presenta como ejemplo el estudio de articulaciones universidad-empresa a partir de la discusión de resultados empíricos derivados de relevamientos hechos en América Latina. Esto implicará la explicitación de patrones de búsqueda específicos (como las diversas denominaciones de la región y la inclusión de los diferentes países que la componen), para ofrecer resultados consistentes que contribuyan a la delimitación precisa de la muestra.

Otro elemento clave a tener en cuenta una vez consolidada la muestra, es el relacionado con la heterogeneidad de los datos. Como se mencionó, los documentos fuente que se utilizan para la construcción del corpus textual en este tipo de análisis no presentan, necesariamente, patrones en común: los artículos de una revista pueden no coincidir en estilo, estructura y diseño con los de otra. Este desafío es usualmente abordado desde la identificación de secciones o porciones de texto que comparten un patrón más allá de su origen de publicación. En el caso de los artículos científicos, el título, el resumen, los autores, la institucionalidad y la bibliografía suelen ser los segmentos más recurrentes al momento de

establecer comparabilidad entre fuentes heterogéneas. Para análisis que vayan más allá de ello, se requerirá la construcción de diccionarios *ad hoc* basados en una exploración manual de los datos, con la finalidad de identificar los límites de los segmentos de texto de interés. Por ejemplo, la sección de antecedentes conceptuales suele presentarse con diferentes denominaciones (marco teórico, antecedentes, revisión conceptual, etc.) y su correcta selección dependerá de la eficiencia en la identificación de estas, así como en el reconocimiento de su finalización.

Desde un punto de vista técnico, el tratamiento de archivos editados con criterios de estructura y estilo similares (por ejemplo, artículos dentro de una misma publicación) facilita enormemente la identificación y estandarización de criterios de selección de porciones de texto. Sin embargo, de trabajar con fuentes heterogéneas, la implementación de un minucioso proceso manual de revisión y caracterización es decisivo para la correcta identificación de secciones y grupos dentro de cada unidad de análisis textual.

Las técnicas frecuentemente utilizadas para este tipo de estudios son de corte tradicional y se basan en el *data mining* y el análisis bibliométrico (Jurowetzki, Lema y Lundvall, 2018). Por supuesto, la minería de texto se apoya sobre una correcta selección de los segmentos dentro de cada artículo, como se señaló en el párrafo precedente. Es habitual que este tipo de estrategias se apoyen sobre el análisis de los *abstract* o resúmenes de los artículos (Evangelopoulos, Prybutok y Zhang, 2012; Dascalu *et al*., 2015) o sobre la bibliografía utilizada combinando técnicas de análisis semántico con análisis bibliométrico (Czerwon, Glänzel y Schubert, 1999; Bonilla, Merigó y Torres-Abad, 2015; Bolici, Deakin y Mora, 2017).

El cuarto ejemplo apunta a destacar algunas de las potencialidades latentes de las técnicas que se plantearon previamente, especialmente en lo que respecta a su complementariedad con técnicas de búsqueda online o *web crawling*. En particular, este caso propone abordar el estudio de documentación estructurada, compuesta por contenido homogéneo en sus secciones, con la particularidad de no estar basado enteramente en la prosa, sino más bien en oraciones cortas y con pautas específicas de estructura. En particular, esta caracterización se aplica en el estudio del currículum vítae (CV) de investigadores latinoamericanos o de otras regiones, aunque también podrían destacarse otras situaciones de utilidad, como el análisis de portadas de artículos (título, resumen, autores, etc.) y el análisis bibliométrico. La principal distinción en el caso de los CV respecto de otros ejemplos, es que su origen puede ser variado, y la

estructura de cada archivo heterogénea. Esto plantea particulares desafíos relacionados con la búsqueda automatizada de información en Internet.

En términos generales, este tipo de aproximación se apoya sobre elementos ya discutidos en los ejemplos previos, aunque se distingue especialmente debido a: 1) los desafíos en torno al *data wrapping* y el *data parsing*; 2) la identificación y codificación de estructuras heterogéneas; 3) la implementación de procesos *ad hoc* de codificación, como respuesta a la heterogeneidad de fuentes, estructuras y formatos.

Para la fase del *web crawling* y *data wrapping*, en este caso será imprescindible contar con un diccionario como referencia genérica para realizar procesos automatizados y eficientes de indagación online. En él, deberán incluirse las diferentes variaciones y combinatorias de criterios de búsqueda involucrados (como variantes del nombre -siglas iniciales, nombre completo-, variantes de la denominación del documento de currículum vítae, etc.). Como se mencionó páginas arriba, el diccionario puede ser utilizado como insumo para paquetes de Python o de R, capaces de realizar un almacenamiento de los principales resultados para su posterior depuración. Existen alternativas en desarrollo, por ejemplo, el buscador SISOB (Barros *et al*., 2015) gratuito y de código abierto, que mediante técnicas de *machine learning* realiza búsquedas automatizadas sobre la base de pocos criterios (nombre, institución y campo de especialización), selecciona documentos con estructura similar a un CV y elabora una base de datos según sus secciones.

El tratamiento y curación de los datos es determinante para evitar el análisis de documentos incompletos, con información faltante, o inadecuados. En este sentido, se sugiere que en la etapa de *parsing* se realicen múltiples chequeos sobre la estructura, para evitar corroboraciones manuales sobre los archivos recolectados. En los casos en los que el corpus lo permita, los chequeos manuales pueden ser más certeros y rápidos que la escritura de diferentes reglas mediante algoritmos.

A partir de la catalogación de este tipo de datos, pueden realizarse análisis sobre las trayectorias laborales, la producción científica, redes de investigadores e institucionales, entre otras. Es usual complementar este tipo de estudios con análisis de redes, ejercicios econométricos y otras técnicas avanzadas de tratamiento de datos que utilizan como insumo las elaboraciones derivadas del análisis semántico.

A modo de síntesis, en la siguiente tabla (ver Cuadro 1) pueden encontrarse las principales características de los casos presentados y las recomendaciones para su aplicación en la práctica.

Cuadro 1. Principales características de los casos presentados

Ejemplo	Fuente de datos	Universo	Consideraciones específicas	Principal desafío	Casos similares
Geopolítica antártica	Registros públicos	Definido	Tratamiento OCR	Especificar claramente niveles y alcance.	Normativas, discusiones parlamentarias, actas, etc.
Pensamiento de CEPAL	Plataforma web específica	Definido	Eficiencia de la database	Tratamiento de datos secuencial y escalonado.	Plataformas indexadas como Scopus, Science Direct, etc.
Relaciones universidad-empresa	Internet/plataformas varias	Difuso	Recorte	Selección de fuentes. Homologación del archivo textual.	Aplicable a múltiples tópicos conceptuales.
Trayectorias científicas en AI	Internet	Difuso	*Web crawling*	Selección de fuentes. Segmentación y validación de secciones de información.	Trayectorias laborales en general, programas académicos, postulaciones a proyectos, etc.

Fuente: elaboración propia

En el cuadro se presenta una síntesis de casos y consideraciones presentadas en esta sección. Como información adicional a lo mencionado líneas arriba, puede destacarse la inclusión de casos similares a los desarrollados en los ejemplos. Una de las cuestiones más relevantes a destacar en la discusión planteada tiene que ver con la posibilidad de trasladar el análisis y los problemas relativos a una temática específica hacia otros casos. Esto habilita amplias posibilidades analíticas y presenta oportunidades aún muy poco exploradas para el estudio de las relaciones de CTIS en América Latina.

Reflexiones en torno a la aplicación de estos métodos en América Latina

En este trabajo se abordaron diversos aspectos metodológicos y conceptuales relacionados con el análisis semántico-estadístico de grandes corpus textuales. En particular, se exploraron los antecedentes y aplicaciones que tienen gran pertinencia para su aplicación en estudios en torno a problemáticas de ciencia, tecnología, innovación y sociedad en América Latina, considerando especialmente la proliferación de información

digital y la gradual disponibilidad en internet de documentos públicos y otros registros de relevancia para la investigación social.

Los métodos y técnicas presentados en este apartado plantean desafíos múltiples, y proponen el foco en la articulación de fenómenos lingüísticos, la informática, la estadística y las ciencias sociales. Así, se señaló que los métodos de análisis semántico-estadísticos son una forma de operar sobre información textual, especialmente aquella presente en entornos digitales, que permiten realizar una interpretación de la realidad que representan. En este marco, se destacó también la importancia de implementar estrategias alineadas con la ciencia abierta, poniendo a disposición pública tanto las herramientas (*scripts*) utilizadas como los datos resultantes de su aplicación, para contribuir a la replicabilidad y la adaptación de estos esfuerzos por parte de la comunidad científica.

Desde aquí, se señalaron los pasos centrales para la implementación de estudios basados en análisis semántico-estadístico, atravesados por el *web crawling* y el *data wrapping*, el *data parsing* y la selección de variables clave, la estructuración de datos y los procesamientos transversales y específicos. Además, se distinguieron herramientas y paquetes de software capaces de facilitar cada una de esas instancias, haciendo énfasis en la disponibilidad de herramientas gratuitas y de código abierto para los lenguajes de programación R y Python. Con dichas herramientas pueden abordarse análisis como la coocurrencia de palabras, estudios sobre el rol de la lingüística computacional en la construcción e interpretación de corpus textuales, y análisis basados en técnicas de procesamiento del lenguaje natural nutridos por el *machine learning*, entre otros.

Las principales reflexiones que se derivan para la aplicación de estas metodologías para problemas de CTIS en América Latina se estructuran sobre tres ejes. El primero es la creciente cantidad de información digital disponible en Internet que no está siendo utilizada con propósitos investigativos; el segundo consiste en que mucha de esta información es generada por instituciones públicas de los países de la región, por lo que se trata de fuentes oficiales, ideales para realizar indagaciones de diversa naturaleza; el tercero es que a la luz de las regulares restricciones presupuestarias en la región y las múltiples limitaciones que pueden existir para la generación *ad hoc* de nuevos datos (por ejemplo, realización de trabajos de campo), los abordajes de análisis semántico constituyen una oportunidad de abordar problemas originales de forma inédita, sobre documentación disponible públicamente.

Lo anterior representa una oportunidad, en especial tomando como punto de partida los ejemplos presentados en el apartado precedente. Todos ellos representan problemas poco explorados y de gran relevancia para la región. El estudio de dichos datos suele no poseer características idóneas, dado que es información que ha sido creada para diversos fines, no necesariamente investigativos. Así, el valor de dichos materiales viene dado por las estrategias de abordaje, depuración y tratamiento que se implementen como parte del proceso de investigación.

El avance en el uso de este tipo de enfoques representa una gran oportunidad para el estudio de problemáticas de ciencia, tecnología, innovación y sociedad en América Latina, ya que permite observar dinámicas intertemporales muchas veces poco evidentes, además de poner en relieve temas y problemas poco explorados. En este sentido, el principal aporte de este trabajo tiene que ver con ofrecer herramientas para poner en valor información original, disponible y accesible, para complementar el estudio de las realidades regionales en nuestros territorios.

Bibliografía

Abdo, A. H.; Cambrosio, A. y Cointet, J.-P. (2020). "Beyond networks: Aligning qualitative and computational science studies". *Quantitative Science Studies*, vol.1, nº 3, pp. 1017-1024. DOI: https://doi.org/10.1162/qss_a_00055.

Adamic, L.; Alstyne, M. V.; Aral, S.; Barabási, A.-L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; Jebara, T.; King, G.; Lazer, D.; Macy, M.; Pentland, A. y Roy, D. (2009). "Computational Social Science". *Science*, vol. 323, nº 5915, pp. 721-723. DOI: https://doi.org/10.1126/science.1167742.

Ain, Q. T.; Ali, M.; Hayat, B.; Kamran, M.; Noureen, A.; Rehman, A. y Riaz, A. (2017). "Sentiment Analysis Using Deep Learning Techniques: A Review". *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 8, nº 6,

Alani, H.; Fernández, M.; He, Y. y Saif, H. (2014). "On stopwords, filtering and data sparsity for sentiment analysis of twitter". En Calzolari, N.; Choukri, K.; Declerck, T.; Loftsson, H.; Maegaard, B.; Mariani, J.; Moreno, A.; Odijk, J. y Piperidis, S. (eds.), *Proceedings of the Ninth International Conference on Language Resources and*

Evaluation (LREC'14), pp. 810-817. Reykjavik, Islandia: European Language Resources Association (ELRA).

Alfonseca, E.; Gliozzo, A.; Magnini, B.; Pérez, D.; Rodríguez, P. y Strapparava, C. (2005). "Sobre los efectos de combinar análisis semántico latente con otras técnicas de procesamiento de lenguaje natural para la evaluación de preguntas abiertas". *Revista signos*, vol. 38, nº 59, pp. 325-343.

Alsadoon, A.; Do, H. H.; Maag, A. y Prasad, P. W. C. (2019). Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review. *Expert Systems with Applications*, vol. 118, pp. 272-299.

Ando, R. K. y Lee, L. (2001). "Iterative residual rescaling". *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 154-162. New York: Association for Computing Machinery.

Araque, O.; Corcuera-Platas, I.; Iglesias, C. A. y Sánchez-Rada, J. F. (2017). "Enhancing deep learning sentiment analysis with ensemble techniques in social applications". *Expert Systems with Applications*, vol. 77, pp. 236-246.

Arza, V. y Fressoli, M. (2017). "Systematizing benefits of open science practices". *Information Services & Use*, vol. 37, nº 4, 463-474. DOI: https://doi.org/10.3233/ISU-170861.

Baker, C. y Fillmore, C. J. (2010). "A frames approach to semantic analysis". En Heine, B. y Narrog, H. (ed.), *The Oxford handbook of linguistic analysis*. Oxford, Reino Unido: Oxford University Press.

Bally, C.; de Saussure, F. y Sechehaye, A. (2003). *Curso de lingüística general*. Buenos Aires: Losada.

Bandaru, N.; Moyer, E. D. y Radhakrishna, S. (2012). *US Patent No. 8/166.013*. Washington, DC: US Patent and Trademark Office.

Bandyopadhyay, S.; Cambria, E.; Das, D. y Feraco, A. (2017). "Affective computing and sentiment analysis". *A practical guide to sentiment analysis*, pp. 1-10. Cham: Springer.

Barbu, A.; Comaniciu, D.; Feulner, J.; Huber, M.; Liu, D.; Seifert, S. y Zhou, S. K. (2009). "Hierarchical parsing and semantic navigation of full body CT data". *Medical Imaging 2009: Image Processing*. SPIE, the international society for optics and photonics.

Barros, B.; Fernandez-Zubieta, A.; Geuna, A.; Guzmán, E.; Kataishi, R.; Lawson, C. y Toselli, M. (2015). "SiSOB data extraction and codification: A tool to analyze scientific careers". *Research Policy*, vol. 44, nº 9, pp. 1645-1658. DOI: https://doi.org/10.1016/j.respol.2015.01.017.

Bauin, S.; Callon, M.; Courtial, J.-P. y Turner, W. A. (1983). "From translations to problematic networks: An introduction to co-word analysis". *Social Science Information*, vol. 22, nº 2, pp. 191-235.

Baya Laffite, N.; Cointet, J.-P.; De Pryck, K.; Gray, I.; Venturini, T. y Zabban, V. (2014). "Three maps and three misunderstandings: A digital mapping of climate diplomacy". *Big Data & Society*, vol. 1, nº 2, pp. 1-21. DOI: https://doi.org/10.1177/2053951714543804.

Bensaude-Vincent, B. (2014). "The politics of buzzwords at the interface of technoscience, market and society: The case of 'public engagement in science'". *Public Understanding of Science*, vol. 23, nº 3, pp. 238-253. DOI: https://doi.org/10.1177/0963662513515371.

Berant, J. y Herzig, J. (2019). "Don't paraphrase, detect! Rapid and Effective Data Collection for Semantic Parsing". ArXiv: 1908.09940. DOI: https://doi.org/10.48550/arXiv.1908.09940.

Berry, D. M. (ed.) (2012). *Understanding digital humanities*. Londres: Palgrave Macmillan.

Blum, B.; Steyvers, M.; Tenenbaum, J. B. y Wagenmakers, E. J. (2003). "Inferring causal networks from observations and interventions". *Cognitive science*, vol. 27, nº 3, pp. 453-489.

Bolici, R.; Deakin, M. y Mora, L. (2017). "The first two decades of smart-city research: A bibliometric analysis". *Journal of Urban Technology*, vol. 24, nº 1, pp. 3-27.

Bonilla, C. A.; Merigó, J. M. y Torres-Abad, C. (2015). "Economics in Latin America: a bibliometric analysis". *Scientometrics*, vol. 105, nº 2, pp. 1239-1252.

Brixner, C.; Lerena, O.; Minervini, M. y Yoguel, G. (2021). "La relación entre la universidad y la empresa: Identificación de comunidades temáticas". *Revista de la CEPAL*, nº 135.

Cai, D.; Han, J. y He, X. (2005). "Document clustering using locality preserving indexing". *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, nº 12, pp. 1624-1637.

Callon, M.; Courtial, J.-P. y Laville, F. (1991). "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry". *Scientometrics*, vol. 22, nº 1, pp. 155-205.

Callon, M.; Law, J. y Rip, A. (eds.) (1986). *Mapping the dynamics of science and technology: Sociology of science in the real world*. Londres: Palgrave Macmillan.

Castro Gouveia, F. y Texeira Rabello, E. (2019). "Métodos digitais nos estudos em saúde. Mapeando usos e propondo sentidos". En Omena, J. J. (ed.), *Métodos Digitais: teoria-prática-crítica*, pp. 136-197. Portugal: Instituto de Comunicação da NOVA.

Collier, N. y Mullen, T. (2004). "Sentiment analysis using support vector machines with diverse information sources". En Lin, D. y Wu, D. (eds.), *Proceedings of the 2004 conference on empirical methods in natural language processing*, pp. 412-418. Barcelona: Association for Computational Linguistics.

Collins, A. M. y Quillian, M. R. (1969). "Retrieval time from semantic memory". *Journal of verbal learning and verbal behavior*, vol. 8, nº 2, pp. 240-247.

Corley, C.; Mihalcea, R. y Strapparava, C. (2006). "Corpus-based and knowledge-based measures of text semantic similarity". En Cohn, A. (ed.), *Proceedings of the 21st national conference on Artificial intelligence*, pp. 775-780. Boston, MA: AAAI Press.

Czerwon, H. J.; Glänzel, W. y Schubert, A. (1999). "A bibliometric analysis of international scientific cooperation of the European Union (1985-1995)". *Scientometrics*, vol. 45, nº 2, pp. 185-202.

Danilevsky, M.; Gu, Q.; Han, J. y Li, Z. (2012). "Locality preserving feature learning". *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics. Proceedings of Machine Learning Research (PMLR)*, vol. 22, pp. 477-485.

Dascalu, M.; Dessus, P.; Paraschiv, I. C. y Trausan-Matu, S. (2015). "Analyzing the Semantic Relatedness of Paper Abstracts: An Application to the Educational Research Field". En Dumitrache, I. (ed.), *2015*

20th International Conference on Control Systems and Computer Science, pp. 759-764. Piscataway, NJ: IEEE.

Delanty, G. y Isin, E. F. (eds.), *Handbook of historical sociology*. Londres: SAGE.

Desrosières, A. (2013). "Les mots et les nombres: Pour une sociologie de l'argument statistique". *Gouverner par les nombres. L'argument statistique II*, pp. 7-35. París: Presses des Mines.

Digital Methods Initiative (2020). "Post-API Research? On the contemporary study of social media data. Digital Methods Winter School and Data Sprint 2020". Disponible en: https://wiki.digitalmethods.net/Dmi/WinterSchool2020.

DiMaggio, P. (2015). "Adapting computational text analysis to social science (and vice versa)". *Big Data & Society*, vol. 2, nº 2. DOI: https://doi.org/10.1177/2053951715602908.

Donoho, D. (2017). "50 Years of Data Science". *Journal of Computational and Graphical Statistics*, vol. 26, nº 4, pp. 745-766. DOI: https://doi.org/10.1080/10618600.2017.1384734.

Dumais, S. T. y Landauer, T. K. (1997). "A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge". *Psychological Review*, vol. 104, nº 2, pp. 211-240. DOI: https://doi.org/10.1037/0033-295X.104.2.211.

Evangelopoulos, N.; Prybutok, V. R. y Zhang, X. (2012). "Latent Semantic Analysis: five methodological recommendations". *European Journal of Information Systems*, vol. 21, nº 1, pp. 70-86.

Fahmy, A. A. y Gomaa, W. H. (2013). "A survey of text similarity approaches". *International Journal of Computer Applications*, vol. 68, nº 13, pp. 13-18.

Farahat, A.; Levow, G.-A.; Matveeva, I. y Royer, C. (2005). "Term representation with generalized latent semantic analysis". En Angelova, G.; Bontcheva, K.; Mitkov, R. y Nicolov, N. (eds.), *Recent Advances in Natural Language Processing IV*, pp. 45-54. Amsterdam: John Benjamins.

Fayyad, U. y Hamutcu, H. (2020). "Toward Foundations for Data Science and Analytics: A Knowledge Framework for Professional Stan-

dards". *Harvard Data Science Review*, vol. 2, nº 2. DOI: https://doi.org/10.1162/99608f92.1a99e67a.

Foltz, P. W.; Kintsch, W. y Landauer, T. K. (1998). "The measurement of textual coherence with latent semantic analysis". *Discourse processes*, vol. 25, nº 2-3, pp. 285-307.

Foltz, P. W.; Kintsch, W.; Laham, D.; Landauer, T. K.; Rehder, B.; Schreiner, M. E. y Wolfe, M. B. (1998). "Learning from text: Matching readers and texts by latent semantic analysis". *Discourse processes*, vol. 25, nº 2-3, pp. 309-336.

Foltz, P. W.; Laham, D. y Landauer, T. K. (1998). "An introduction to latent semantic analysis". *Discourse processes*, vol. 25, nº 2-3, pp. 259-284.

Foucault, M. (2017). *La arqueología del saber*. Buenos Aires: Siglo XXI.

Geeraerts, D.; Heylen, K.; Speelman, D. y Wielfaert, T. (2015). "Monitoring polysemy: Word space models as a tool for large-scale lexical semantic analysis". *Lingua*, vol. 157, pp. 153-172.

Ghose, A.; Ipeirotis, P. y Sundararajan, A. (2007). "Opinion mining using econometrics: A case study on reputation systems". En van den Bosch, A. y Zaenen, A. (eds.), *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 416-423. Praga: Association for Computational Linguistics.

Golder, S. A. y Macy, M. W. (2014). "Digital Footprints: Opportunities and Challenges for Online Social Research". *Annual Review of Sociology*, vol. 40, nº 1, pp. 129-152. DOI: https://doi.org/10.1146/annurev-soc-071913-043145.

Griffiths, T. L. y Steyvers, M. (2002). "A probabilistic approach to semantic representation". En Gray, W. D. y Schunn, C. D. (eds.), *Proceedings of the annual meeting of the cognitive science society*. Londres: Routledge.

Grimmer, J. y Stewart, B. M. (2013). "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts". *Political Analysis*, vol. 21, nº 3, pp. 267-297. DOI: https://doi.org/10.1093/pan/mps028.

Grimmer, J.; Roberts, M. E. y Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton: Princeton University Press.

Hampton, K. N. (2017). "Studying the Digital: Directions and Challenges for Digital Methods". *Annual Review of Sociology*, vol. 43, nº 1, pp. 167-188. DOI: https://doi.org/10.1146/annurev-soc-060116-053505.

Hardie, A. y McEnery, T. (2013). "The History of Corpus Linguistics". En Allan, K. (ed.), *The Oxford Handbook of the History of Linguistics*. Oxford, Reino Unido: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199585847.013.0034.

Hearst, M. (2003). "What is text mining?". Disponible en: https://people.ischool.berkeley.edu/~hearst/text-mining.html.

Heyes La Bond, C.; Lupton, D.; Pink, S. y Sumartojo, S. (2017). "Mundane data: The routines, contingencies and accomplishments of digital living". *Big Data & Society*, vol. 4, nº 1. DOI: https://doi.org/10.1177/2053951717700924.

Hilbert, M. y López, P. (2011). "The world's technological capacity to store, communicate, and compute information". *Science*, vol. 332, nº 6025, pp. 60-65.

Hirst, G. (2013). "Computational Linguistics". En Allan, K. (ed.), *The Oxford Handbook of the History of Linguistics*. Oxford, Reino Unido: Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199585847.013.0033.

Hockey, S. (2004). "The History of Humanities Computing". En Schreibman, S.; Siemens, R. y Unsworth, J. (eds.), *A Companion to Digital Humanities*, pp. 1-19. New York: John Wiley & Sons. DOI: https://doi.org/10.1002/9780470999875.ch1.

Hofmann, T. (2013). "Probabilistic latent semantic analysis". En Laskey, K. B. y Prade, H. (eds.), *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. San Francisco, CA: Morgan Kaufmann. ArXiv: 1301.6705.

Inkpen, D. e Islam, A. (2008). "Semantic text similarity using corpus-based word similarity and string similarity". *ACM Transactions on Knowledge Discovery from Data*, vol. 2, nº 2, pp. 1-25. DOI: https://doi.org/10.1145/1376815.1376819.

Jurowetzki, R.; Lema, R. y Lundvall, B.-Å. (2018). "Combining Innovation Systems and Global Value Chains for Development: Towards a Research Agenda". *The European Journal of Development Re-*

search, nº 30, pp. 364-388. DOI: https://doi.org/10.1057/s41287-018-0137-4.

Kintsch, W. (2002). "The potential of latent semantic analysis for machine grading of clinical case summaries". *Journal of biomedical informatics*, vol. 35, nº 1, pp. 3-7.

Koselleck, R. (2004). *Futures past: On the semantics of historical time*. New York: Columbia University Press.

Kreimer, P. y Zuckerfeld, M. (2014). "La explotación cognitiva: tensiones emergentes en la producción y uso social de conocimientos científicos, tradicionales, informacionales y laborales". En Arellano, A.; Kreimer, P.; Velho, L. y Vessuri, H. (coords.), *Perspectivas latinoamericanas en el estudio social de la ciencia, la tecnología y la sociedad* (pp. 178-193). España: Siglo XXI.

Lazer, D. y Radford, J. (2017). "Data ex Machina: Introduction to Big Data". *Annual Review of Sociology*, 43(1), pp. 19-39. DOI: https://doi.org/10.1146/annurev-soc-060116-053457.

Lieber, E. (2009). "Mixing qualitative and quantitative methods: Insights into design and analysis issues". *Journal of Ethnographic & Qualitative Research*, vol. 3, nº 4.

Liu, B.; Wang, S. y Zhang, L. (2018). "Deep learning for sentiment analysis: A survey". *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 8, nº 4.

Lyon, L. (2016). "Transparency: The emerging third dimension of Open Science and Open Data". *LIBER Quarterly*, vol. 25, nº 4, pp. 153-171. DOI: https://doi.org/10.18352/lq.10113.

Marres, N. y Weltevrede, E. (2013). "Scraping the Social?". *Journal of Cultural Economy*, vol. 6, nº 3, pp. 313-335. DOI: https://doi.org/10.1080/17530350.2013.772070.

Moretti, F. (2013). *Distant reading*. New York/Londres: Verso Books.

Peine, A.; Spitters, C. y van Lente, H. (2013). "Comparing technological hype cycles: Towards a theory". *Technological Forecasting and Social Change*, vol. 80, nº 8, pp. 1615-1628. DOI: https://doi.org/10.1016/j.techfore.2012.12.004.

Peirsman, Y. (2008). "Word space models of semantic similarity and relatedness". En Balogh, K. (ed.), *Proceedings of the 13th ESSLLI*

Student Session, pp. 143-152. Amsterdam: Institute for Logic, Language and Computation (ILLC) of the Universiteit van Amsterdam.

Rip, A. y Voß, J.-P. (2013). "Umbrella terms as mediators in the governance of emerging science and technology". *Science, technology and innovation studies: STI studies*, vol. 9, n° 2, pp. 39-59.

Rogers, R. (2013). *Digital methods*. Massachusetts: MIT Press.

Ruppert, E. y Scheel, S. (2019). "The Politics of Method: Taming the New, Making Data Official". *International Political Sociology*, vol. 13, n° 3, pp. 1-20. DOI: https://doi.org/10.1093/ips/olz009.

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton: Princeton University Press.

Skinner, Q. (1969). "Meaning and understanding in the history of ideas". *Visions of politics*. Cambridge: Cambridge University Press.

Strogatz, S. H. y Watts, D. J. (1998). "Collective dynamics of 'small-world' networks". *Nature*, vol. 393, n° 6684, pp. 440-442.

Vera, P. (2016). "Imaginarios urbanos tecnológicos: los hilos de las construcciones sociotécnicas de la ciudad". *Horizontes Sociológicos*, n° 8, pp. 147-164.

Watts, D. J. (2011). *Everything is obvious: Once you know the answer*. New York: Crown Business.

Capítulo 8
Modelos estructurales cualitativos para el estudio y comprensión de los procesos de ciencia, tecnología e innovación

Mayela Saraí López-Castro, Nayeli Martínez, Natalia Gras, José Miguel Natera

Introducción

Al abordar temáticas asociadas al análisis y comprensión de los fenómenos sociales, como los procesos de ciencia, tecnología e innovación (CTI), surgen problemáticas relacionadas con el tipo de información existente y disponible. En particular, los países en desarrollo enfrentan un desafío singular respecto a la calidad y suficiencia de información sistematizada en grandes bases de datos homogéneas.

Surgen así preguntas sobre qué métodos y herramientas de investigación permiten analizar los procesos de CTI y extraer conclusiones robustas, teniendo en consideración dos grandes cuestiones. La primera es que los fenómenos sociales son complejos y multidimensionales e implican una serie de dificultades asociadas con la presencia de un gran número de elementos (hechos, actores, acciones, características, etc.) que están directa o indirectamente relacionados. Esta complejidad dificulta el desarrollo de sistemas de medición de la CTI de calidad, especialmente en los países de América Latina y el Caribe (ALC), donde las deficiencias informacionales se agudizan. A lo que se suma la segunda cuestión, la cual está vinculada a la falta o insuficiente información cuantitativa para observar las especificidades de dichos fenómenos.

Así, las aproximaciones metodológicas cualitativas robustas se vuelven relevantes y necesarias pues, por su flexibilidad, permiten el análisis de fenómenos sociales que pueden verse limitados por la rigidez que pueden imponer los métodos cuantitativos. Lo anterior facilita el "realismo" en el estudio observacional de fenómenos complejos y las posibilidades de inferencia, respondiendo a preguntas sobre el cómo, el cuándo y el porqué de los eventos.

El objetivo de este trabajo es mostrar cómo los modelos estructurales cualitativos (en adelante, MEC) proveen una herramienta válida para superar las dificultades mencionadas en el estudio de los procesos de CTI. Los MEC son métodos de análisis diseñados para deducir e interpretar estructuras, dimensionalidad y relaciones subyacentes en los fenómenos complejos; utilizan información cualitativa y la operacionalizan para dar una mayor claridad tanto en la estrategia de investigación seguida como en las características de las estructuras analizadas. En otras palabras, ofrecen la posibilidad de transformar "modelos mentales poco claros y mal articulados, en modelos visibles y bien definidos" (Attri, Dev y Sharma, 2013: 3). Es por ello que consideramos que tienen el potencial para avanzar considerablemente en la comprensión, análisis y explicación de los procesos de CTI, particularmente en ALC.

En este trabajo presentamos una alternativa metodológica basada en los MEC, particularmente en dos métodos: 1) el modelado estructural interpretativo total (MEIT) y 2) el análisis estructural causal cualitativo (AECC). Si bien ambos métodos son estructurales, tienen enfoques distintos. El MEIT analiza problemas muy específicos permitiendo descubrir y modelar diversos tipos de relaciones entre variables homogéneas, por ejemplo, conocer la causa, problemas, oportunidades, etc., en un problema determinado; y su método es muy estructurado. El AECC, por su parte, analiza problemas o realidades desconocidas permitiendo descubrir y modelar relaciones causa-efecto entre variables heterogéneas y su método es más flexible.

Después de esta introducción, se muestra la utilidad de los modelos estructurales ante los desafíos para la investigación de corte cualitativo de procesos de CTI. El tercer apartado describe los dos métodos propuestos a través de ejemplos ilustrativos de su aplicación en estudios de CTI. Luego se presenta la discusión sobre las ventajas y limitaciones de ambos métodos. Finalmente, en el quinto apartado se presentan las conclusiones.

Utilidad de los modelos estructurales ante los desafíos para la investigación cualitativa

El estudio de fenómenos sociales complejos tales como los procesos de CTI implica problemas de investigación que generalmente requieren información de tipo cualitativa. En esta situación, los métodos estadísticos tienen limitaciones para capturar la complejidad que encierra el análisis de estos fenómenos (Pearl, 2009). De aquí la relevancia de utilizar diseños alternativos que permitan analizar problemas sociales desde distintos ángulos.

La investigación cualitativa ha estado dominada por los estudios de caso (únicos o múltiples), por la etnografía y la fenomenología. Estos métodos proveen diversas técnicas para la recolección de información y generación de hipótesis, así como herramientas para analizar y validar los resultados, observación directa, la documentación, la triangulación, las entrevistas semiestructuradas y el análisis de narrativas (ver, Glaser y Strauss, 1967; Strauss, 1987). Sin embargo, existen otros métodos que han sido poco utilizados en este tipo de investigación, pero que han sido generalmente aplicados a investigaciones experimentales de las ciencias naturales y exactas, y que proveen ventajas para el estudio y la comprensión de los fenómenos sociales complejos y multidimensionales (Maxwell, 2004; Pearl, 2009).

Los MEC se basan en la propuesta del realismo crítico como alternativa para las ciencias sociales frente a la definición y utilización del concepto de causalidad de los paradigmas positivista e interpretativo[12] (Barco y Carrasco, 2018); pues plantea que, al conceptualizar y contextualizar los eventos e identificar los mecanismos que los generan, es posible realizar explicaciones causales utilizando estrategias y análisis para extraer e interpretar relaciones causales de los acontecimientos sociales (Ruffa y Evangelista, 2021; Maxwell, 2021).

Los MEC contribuyen al análisis riguroso de información cualitativa sobre las relaciones entre los distintos elementos (hechos, actores,

12 Mientras el *paradigma positivista* considera la causalidad como un resultado experimental con regularidad probabilística, el *paradigma interpretativo* niega el concepto de causalidad para estudiar los fenómenos sociales, al considerar que causa y efecto se producen de forma simultánea. De acuerdo con Salmon (1984) esta tensión generó el avance de la metodología cuantitativa y cualitativa de forma separada, en la llamada "guerra de paradigmas". El realismo crítico (radical) busca romper con la hegemonía de la causalidad positivista y superar una guerra que, según este autor, ha limitado la investigación cualitativa.

políticas, etc.) subyacentes al problema complejo de investigación. Estos métodos permiten descubrir la estructura implícita y poco clara entre esos elementos (Salmon, 1984; von Wright, 1987; Pearl, 2009). Las herramientas utilizadas en los MEC (matrices estructurales, dígrafos[13] y diagramas causales) permiten un tratamiento riguroso de la información cualitativa, considerando las limitaciones derivadas de la subjetividad inherente al análisis de quien investiga, presente en cualquier otro método cualitativo o cuantitativo (Pearl, 2009; Attri, Dev y Sharma, 2013; Barco y Carrasco, 2018; Blersch *et al*. 2021). Este tipo específico de modelos permite –en comparación con otros estudios observacionales– una mayor abstracción del problema complejo de investigación, la identificación de sus relaciones estructurales y, con ello, una mayor claridad sobre los resultados y hallazgos obtenidos. Los MEC habilitan un *análisis inductivo-deductivo* para la generación de nuevas categorías conceptuales y sus relaciones. Desde la evidencia (inducción) y con un fundamento teórico (deducción) se construyen, identifican y validan categorías conceptuales y sus relaciones, que se contrastan con la teoría hasta ese momento existente. Es decir, sin hacer uso de técnicas estadísticas, parten de la contextualización, abstracción y estilización de la información (Burt, 1982; Sushil, 2012; Maxwell, 2021).

El fundamento de los MEC, y su principal diferencia con los estudios de caso, es que permiten agrupar relaciones entre un gran número de variables en unos pocos factores de análisis (Pearl, 2009). Dichas agrupaciones se basan en la teoría y el conjunto de elementos y relaciones que caracterizan una determinada situación o estructura subyacente a la información cualitativa generada y sistematizada (Maxwell, 2004). Por lo anterior y desde que se introdujeron los MEC para analizar problemas sociales complejos, estudios recientes han planteado el uso potencial de estos métodos en los distintos campos de estudio de las ciencias sociales (Sushil, 2018; Menon y Suresh, 2020). Sin embargo, los estudios de CTI basados en MEC aún son escasos, dejando un campo fértil para la investigación aplicada. El cuadro 1 muestra algunos trabajos relacionados con la CTI que han aplicado los MEC.

13 Es un término derivado de *directed graph* y, como su nombre lo indica, es una representación gráfica de las variables, sus relaciones y los niveles jerárquicos (Attri, Dev y Sharma, 2013).

Cuadro 1. Investigaciones en CTI que han utilizado MEC

Autor y año	Problema	Método
Akriti, Sharma y Vigneswara (2018)	Factores para medir la innovación en las universidades	MEIT
Haleem, Kumar y Luthra (2018)	Gestión de la innovación de productos	MEIT
Haleem, Husain y Khan (2017)	Barreras a la transferencia de tecnología	MEIT
Fatma *et al.* (2022)	Barreras hacia la adopción de la iniciativa empresarial estratégica	MEIT
Kumbhat y Sushil (2022)	Interacciones entre los factores de éxito de los *startups* de alta tecnología	MEIT
López-Castro (2019)	Interacciones entre los obstáculos a la transferencia de conocimiento entre las universidades y empresas y los factores organizacionales	MEIT
Dutrénit *et al.* (2018)	Relaciones entre actores y hechos clave para el desarrollo de innovaciones inclusivas	AECC
Mukunda Das *et al.* (2019)	Interacciones entre diferentes factores institucionales y ecosistemas empresariales	MEIT
Dhir y Singh (2021)	Modelización de los antecedentes de la aplicación de la innovación	MEIT
Dutrénit, Gras y Vera-Cruz (2019)	Factores que determinan la creación, el éxito y el impacto de innovaciones con fines sociales	AECC

Fuente: elaboración propia

Propuesta metodológica: dos modelos estructurales alternativos para la investigación cualitativa de los procesos de CTI

En esta sección se presentan los dos métodos propuestos en este trabajo: el modelado estructural interpretativo total (MEIT) y el análisis estructural-causal cualitativo (AECC). La descripción de cada uno de ellos se acompaña con ejemplos, a efectos de proporcionar una mayor claridad en la aplicación a problemas reales.

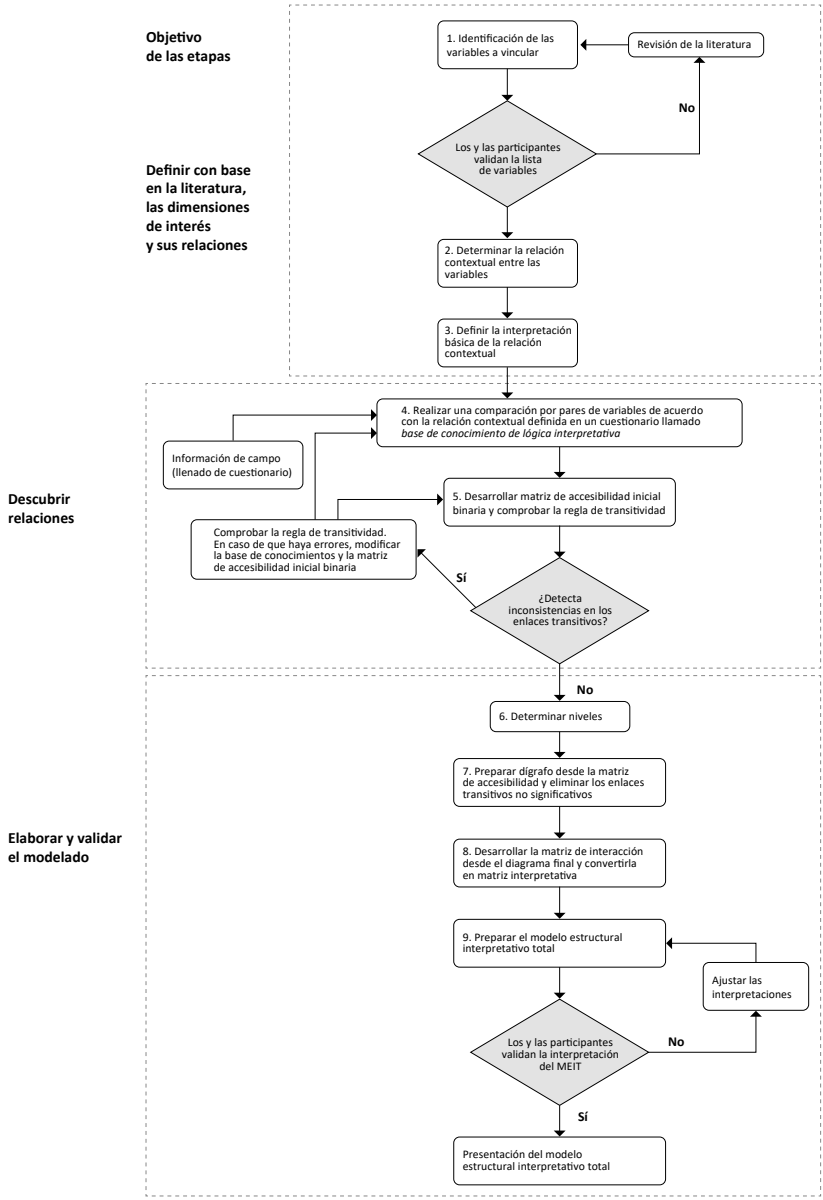
El modelado estructural interpretativo total (MEIT) y su evolución

El MEIT (en inglés, Total Interpretative Structural Modelling (TISM)) es una evolución y mejora de su antecesor: el modelado estructural interpretativo (MEI) (en inglés, Interpretative Structural Modelling (ISM)), técnica introducida por Warfield (1974) que permite identificar la relación entre las variables consideradas en un problema específico, percibir mejor la estructura del sistema a través de una representación visual del problema y desarrollar una jerarquía en función de la importancia de las variables (Faisal, 2010).

El MEIT (Sushil, 2012), en esencia, es lo mismo que el MEI, sin embargo, la diferencia más significativa es que el MEIT interpreta la relación entre los elementos –causalidad– mientras que el MEI no (Jena *et al*., 2017; Menon y Suresh, 2020). El MEI, por ejemplo, solo ayuda a proporcionar respuestas a "qué" y "cómo" en la construcción de teorías. Sin embargo, permanece en silencio sobre la causalidad de los vínculos y, por lo tanto, no puede responder "por qué". El MEIT, por su parte, facilita las respuestas por el "qué", "cómo" y "por qué" en la construcción de teorías y es por este hecho que el MEIT supera el principal inconveniente de su antecesor, es decir, la nula interpretación de los enlaces.

Dadas las ventajas observadas del MEIT, a continuación, se exponen los pasos básicos para desarrollar y aplicar la técnica en un problema de investigación, con la ayuda de un diagrama esquemático como se muestra en el diagrama 1.

Diagrama 1. Pasos para el desarrollo y aplicación del modelado estructural interpretativo total



Fuente: adaptado de Sushil (2012) y Jena *et al.*, (2017)

Con el fin de mostrar este método mediante un ejemplo aplicado en el campo de la CTI, se partirá del trabajo de López-Castro (2019); en él se muestran las relaciones entre la estructura organizacional del Tecnológico Nacional de México en Celaya y los obstáculos en la transferencia de conocimiento con las empresas de la región.

Paso 1. Identificación de variables

El primer paso comienza con la identificación de variables que son relevantes para el problema de investigación. Estas variables se definen con base en la revisión de la literatura y en los aportes de los y las participantes del campo (a quienes se puede acceder a través de grupos focales, entrevistas, cuestionarios, etc.).

Antes de iniciar el proceso del modelado se debe responder una pregunta crítica: ¿se incluyen todas las variables relevantes?

La solución a esta pregunta está en la revisión de la literatura (deducción) y la interacción con el campo (inducción). Se deberá realizar una revisión estructurada de la lista de variables de la literatura para asegurarse de que se incluyen todas las variables relevantes al problema de estudio. Seguidamente, de forma iterativa, en esta etapa también se debe realizar la verificación de las variables en el contexto de la investigación. Este vaivén entre la literatura y el campo se realiza para comprobar que la lista de variables iniciales es relevante en el contexto de análisis. Lo anterior es imprescindible porque se puede dar el caso de que algunas variables identificadas en la literatura pueden no ser relevantes en el contexto de análisis o, por el contrario, pueden emerger otras variables que no estaban consideradas. Es de suma importancia, por lo tanto, validar tanto teórica como empíricamente la lista de variables a utilizar en el modelo antes de continuar con la aplicación de la técnica.

En el trabajo de López-Castro (2019), la estructura organizativa y los obstáculos se organizan de acuerdo con la lista de variables que se muestran en el cuadro 2. Esta lista de variables se extrajo de una revisión de la literatura y fue verificada con los y las participantes del campo del Tecnológico Nacional de México en Celaya.

Cuadro 2. Variables de la estructura organizativa y los obstáculos en los procesos de transferencia de conocimiento determinadas por la validación de los y las participantes del campo

Número de variable	Dimensión	Variable
V1	**Estructura formal**	Formalización (adaptada y basada en la experiencia)[1]
V2		Descentralización[2]
V3		Complejidad (especialización de miembros basadas en la experiencia)[1]
V4	**Estructura informal**	Comunicación intraorganizacional
V5		Confianza intraorganizacional
V6		Colaboración intraorganizacional
V7	**Obstáculos**	Falta de procedimientos establecidos para la colaboración con la industria
V8		Complejidad y lentitud en los procesos
V9		Falta de personal que apoye las tareas
V10		Falta de profesionalización del personal universitario, especialmente los que gestionan las relaciones
V11		Rigidez de las reglas y regulaciones impuestas por las IES o agencias de financiamiento del gobierno

Notas: (1) Modificada y adaptada al contexto; (2) la variable de centralización planteada en un inicio se ha manifestado inversamente, es decir, lo que se manifiesta en el contexto de estudio es descentralización, por lo cual se tomó como tal, siendo una dimensión propiamente de la estructura organizacional.

Fuente: adaptado de López-Castro (2019)

Paso 2. Determinar la relación contextual entre las variables

El objetivo de este paso es permitirle a quien investiga definir la relación contextual entre las variables identificadas. Para este propósito, MEIT ofrece varias alternativas para establecer la relación contextual entre las variables que se van a analizar (ver cuadro 3).

Cuadro 3. Estructuras de relación contextual realizadas en el MEIT

Tipo	Explicación		
	Elementos/variables	**Relación**	**Interpretación**
Intención	Objetivos	A ayudará a lograr B	¿de qué manera A ayudará a lograr B?
Prioridad, clasificación	Proyectos, metas, etc.	A es de igual o mayor prioridad que B	¿sobre qué base se decide la prioridad?
Mejora de atributo	Problemas, oportunidades, causas	A influiría, mejoraría o causaría B	¿cómo A podría influir, mejorar o causar B?
Estructura del proceso	Actividades, eventos, etc.	A debe preceder a B	¿por qué debería A preceder a B?
Dependencia matemática	Parámetros o factores cuantificables	A es una función de B	¿cuál es la naturaleza de la función entre A y B?

Fuente: adaptado de Sushil (2012: 92)

Es fundamental que quien investiga tenga claro, antes de llegar a los y las participantes de campo, qué tipo de relación se va establecer entre las variables para abordar el problema. Esto asegura que la relación entre variables esté planteada de la manera correcta. Por ejemplo, si quien investiga quiere identificar posibles problemas, oportunidades o causas entre variables, deberá escoger el tipo de estructura mejora de atributo estableciendo la relación de acuerdo con las alternativas según sea el caso (A influye en B, A mejora B o A causa B); o bien, si desea conocer la importancia de variables en un problema de investigación, deberá escoger el tipo de estructura prioridad o clasificación (verificando si A es de igual o mayor prioridad que B). Se recomienda precisar un solo tipo de relación en el enunciado para evitar ambigüedades o confusión en el planteamiento, pues si bien este método permite integrar varios tipos de estructuras de relación en el problema, la mezcla de ellos genera un aumento exponencial en la complejidad durante la aplicación.

En el caso que presentamos (López-Castro, 2019), se tomó como base el tipo de estructura mejora de atributo, pues se trata de un problema que se quiere resolver. El objetivo del trabajo incluye generar recomendaciones de políticas que permitan incentivar y dinamizar el proceso de transferencia de conocimiento desde la academia hacia la industria.

Paso 3. Interpretación de la relación

Una vez que se ha determinado el tipo de relación entre variables, quien investiga debe definir la interpretación de la relación tal y como

se muestra en el cuadro 3. Esto permite comprender cómo opera la relación direccional en el sistema que se está considerando. Retomando el ejemplo anterior, en el primer caso los y las participantes del campo deberán responder una de las tres preguntas según sea el caso: ¿cómo A influye en B?, ¿cómo A mejora B? o ¿cómo A causa B? Por el contrario, en el segundo caso se debe responder: ¿sobre qué base se decide la prioridad?

En el caso que presentamos (López-Castro, 2019), el tipo de relación contextual se interpreta haciendo dos planteamientos: 1) la relación desde las variables de la estructura organizacional hacia las demás variables se analizó considerando la posibilidad de mejora y 2) la relación desde los obstáculos hacia las demás variables se analizó considerando la posibilidad de influencia.

Paso 4. Lógica interpretativa de la comparación por pares de variables

El objetivo de este paso es sistematizar y ordenar la información de los pasos previos a la vista al campo. Para ello, se propone crear un cuestionario comúnmente conocido como base de conocimiento de lógica interpretativa, que reflejará de manera ordenada y clara la relación entre las variables identificadas, así como su interpretación. En este cuestionario se realizará una comparación por pares de variables, es decir, la variable V1 se compara individualmente con todas las otras variables, menos consigo misma. La respuesta para cada comparación puede ser SÍ o NO. Si la respuesta es SÍ, es necesaria una interpretación adicional.

Este paso lo ilustramos con el caso estudiado por López-Castro (2019): las variables del cuadro 2 han sido analizadas bajo el tipo de relación contextual de mejora de atributo, interpretándose desde la posibilidad de mejora o de influencia (según sea el caso). Lo anterior implica generar una base de conocimiento con $q = 11$ variables (seis de estructura organizacional y cinco obstáculos). El cuadro 4 muestra parte del cuestionario empleado, el cual es la base de conocimiento de lógica interpretativa y, por razones de espacio, no se muestra de forma completa, pues implica el análisis de ciento diez relaciones entre variables. Este cuestionario fue adaptado en su propuesta original (ver, Sushil, 2020: 98) y elaborado con información de los y las participantes de campo a quienes se pudo acceder, a través de grupos focales y entrevistas. Fue muy importante seleccionar a participantes con alto nivel de implicación en el proceso de transferencia de conocimiento y tecnología en la academia (gestión, generación de investigación y desarrollos tecnológicos).

Cuadro 4. Base de conocimiento de lógicas interpretativas

Número de comparaciones[a]	Número de variable	Comparación por pares de variables	SÍ/NO	¿Cómo o de qué manera la variable V1 mejora?
	V1	**Las normativas existentes en su instituto tecnológico referidas a los contratos con las empresas en su organización MEJORAN…**		
1	V1-V2	… la descentralización de su organización.	SÍ	Distribución de poder y responsabilidades, mecanismo de motivación para involucrados (libertad en la toma decisiones), agilizar procesos
2	V1-V3	… la especialización que los miembros de su organización tienen en la actualidad.	SÍ	Nuevas responsabilidades, más especialización
3	V1-V4	… la comunicación entre el personal de su organización.	SÍ	Canales abiertos, informales y rápidos para responder a las necesidades de interesados
4	V1-V5	… la confianza que tiene en su organización.	SÍ	Creación de normas *ad hoc* para incentivar y respaldar la colaboración con empresas, el liderazgo directivo y la reputación local y nacional
5	V1-V6	… la colaboración de los miembros de su organización.	SÍ	Autoorganización, incentivos no monetarios para fomentar la colaboración interna
6	V1-V7	… los procedimientos para la colaboración con la industria en su instituto tecnológico.	SÍ	Reducir la burocracia en los procesos
7	V1-V8	… la complejidad y lentitud en los procesos.	SÍ	Atribución de responsabilidades al CIIT[b] en procesos de vinculación (órgano descentralizado responsable de la gestión con las empresas)
8	V1-V9	… la falta de personal que apoye las tareas de vinculación.	NO	-
9	V1-V10	… la falta de profesionalización de los miembros de su organización.	SÍ	Nuevas responsabilidades, más especialización
10	V1-V11	… la rigidez de reglas impuestas por su Institución o las agencias de financiamiento del gobierno.	NO	-
Número de comparaciones	**Número de variable**	**Comparación por pares de variables**	**SÍ/NO**	**¿Cómo o de qué manera la variable V7 influye?**
	V7	**La falta de procedimientos para la colaboración con la industria en su instituto tecnológico INFLUYE en…**		
71	V7-V1	… la descentralización de su organización.	NO	-
72	V7-V2	… las normativas existentes en su instituto tecnológico referidas a los contratos con las empresas.	NO	-
73	V7-V3	… la especialización que los miembros de su organización tienen en la actualidad.	NO	-
74	V7-V4	… la comunicación entre el personal de su organización.	NO	-
75	V7-V5	… la confianza que tiene en su organización.	NO	-
76	V7-V6	… la colaboración de los miembros de su organización.	NO	-
77	V7-V8	… la complejidad y lentitud en los procesos.	SÍ	Necesidad de una búsqueda informal que ralentiza los procesos.
78	V7-V9	… la falta de personal que apoye las tareas de vinculación.	SÍ	Carencia de directrices que permitan estructurar el personal, presupuesto bajo y dependencia para contratar.
79	V7-V10	… la falta de profesionalización de los miembros de su organización.	SÍ	Ausencia de guías y planes de formación a nivel sistema; a nivel interno se han ido creando.
80	V7-V11	… la rigidez de reglas impuestas por su institución o las agencias de financiamiento del gobierno.	NO	-

Notas: (a) Para comprobar el número de comparaciones entre las q variables, se puede utilizar la fórmula $q^2 - q$. En este caso $q = 11$, por lo que: $11^2 - 11 = 110$, en el documento de López-Castro (2019) se muestra el cuadro completo; (b) Centro de Investigación e Innovación Tecnológica al servicio del Tecnológico Nacional de México en Celaya.

Fuente: adaptado de López-Castro (2019)

Paso 5. Matriz de accesibilidad y comprobación de transitividad

En este paso se reestablece el contacto con el campo. El objetivo es codificar las respuestas obtenidas en el cuestionario que deriva del paso anterior. Para ello, las respuestas obtenidas se codifican en una matriz de accesibilidad binaria haciendo la entrada 1 en la celda *i-j*, si la entrada correspondiente en el cuestionario es SÍ, de lo contrario, se debe ingresar como 0 si la respuesta es NO. Es importante tener en cuenta que en las celdas diagonales (que corresponden a la comparación de la variable contra sí misma) se debe insertar un 1.[14]

Respecto al llenado del cuestionario y la matriz binaria, existen dos alternativas. La primera alternativa consiste en que los y las participantes llenen el cuestionario de manera individual y, en este caso, se toma un criterio de mayoría calificada: si más de dos tercios de las y los encuestados declaran una relación como SÍ, debe tratarse como 1, de lo contrario, debe tratarse como 0. La segunda alternativa consiste en que los y las participantes de campo llenen el cuestionario en consenso (por ejemplo, a través de grupos focales) y, en esta situación, se procede directamente al llenado de la matriz de accesibilidad inicial.

Continuando con el caso, López-Castro, en su trabajo utilizó la primera alternativa, por lo que debió construir primero la tabla 1(a), registrando el número de respuestas individuales y determinando si cumplían con el criterio de validación. En ese caso n = 8, por lo que las respuestas mayores 5 cumplen con la mayoría de dos tercios, asignando un 1 en la celda correspondiente, y 0 en el caso contrario; posteriormente, se construyó la tabla 1(b).

14 Esto corresponde a un criterio de coherencia interna. Literalmente, bajo una lógica causal, el "1" significaría que la variable se causa a sí misma; sin embargo, en una perspectiva más razonada, diríamos solo que la variable es pertinente para el estudio.

Tabla 1. (a) Matriz agregada con número de respuestas SÍ en varias celdas (n = 8). (b) Matriz de accesibilidad (criterio 2/3 de mayoría da la entrada 1)

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
(a) Matriz agregada individual											
V1	-	*7*	*6*	*6*	*7*	*6*	*7*	*7*	4	*6*	1
V2	*6*	-	*6*	*7*	*7*	*7*	*6*	*6*	2	*6*	1
V3	4	5	-	*6*	*6*	*6*	*7*	5	*6*	*7*	0
V4	3	4	5	-	*7*	*7*	*6*	*6*	7	*6*	2
V5	5	5	5	*7*	-	*6*	*7*	3	2	*7*	0
V6	5	4	1	5	4	-	*7*	*8*	5	5	0
V7	4	3	5	2	4	5	-	*8*	*6*	*7*	2
V8	5	2	5	3	5	3	4	-	2	5	1
V9	2	1	4	0	5	4	4	*8*	-	5	0
V10	5	2	5	1	4	5	5	*6*	5	-	0
V11	*6*	2	2	1	2	2	3	*8*	4	5	-
(b) Matriz de accesibilidad binaria											
V1	1	1	1	1	1	1	1	1	0	1	0
V2	0	1	1	1	1	1	1	1	0	1	0
V3	0	0	1	1	1	1	1	0	1	1	0
V4	0	0	0	1	1	1	1	1	1	1	0
V5	0	0	0	1	1	1	1	0	0	1	0
V6	0	0	0	0	0	1	1	1	0	0	0
V7	0	0	0	0	0	0	1	1	1	1	0
V8	0	0	0	0	0	0	0	1	0	0	0
V9	0	0	0	0	0	0	0	1	1	0	0
V10	0	0	0	0	0	0	0	1	0	1	0
V11	1	0	0	0	0	0	0	1	0	0	1

Notas: Las cursivas indican las respuestas individuales que cumplen con el criterio de mayoría. Las entradas diagonales son siempre 1 ya que cualquier elemento llega siempre a sí mismo.

Fuente: adaptado de López-Castro (2019)

Después de la codificación binaria, la matriz de accesibilidad obtenida se revisa para determinar la regla de transitividad, en la que se verifica la existencia de relaciones indirectas entre variables. La transitividad establece relaciones entre dos variables como consecuencia de la relación que estas tienen con (al menos) una tercera variable; así, es común que no se evidencie una relación directa entre dos variables en la matriz de accesibilidad y, sin embargo, sí se observe una relación indirecta entre ellas por efecto de la transitividad: es por ello que se realiza una recodificación de la matriz de accesibilidad, con el fin de dar cuenta de estas relaciones indirectas (transitivas). Por ejemplo, en el caso de análisis, López-Castro encontró que la variable V1 estaba directamente relacionada con las variables V2, V3, V4, V5, V6, V7 y V8, entonces, V1 está (al menos) indirectamente relacionada con V9, luego, si en la matriz de accesibilidad inicial, la relación directa de V1 a V9 es 0, debe recodificarse como 1*: en este caso hablaremos de transitividad de primer orden. Para ver una transitividad de segundo orden, veamos el ejemplo de la variable V5 que se relacionó con V6 y V7, entonces la relación V5 V8 marcada por

0 se recodifica a 1* y, por consiguiente, la relación V5 V9 marcada con 0 se recodifica a 1*.

Estos casos de recodificación por la regla de transitividad se muestran en la tabla 2 para el caso ilustrativo. Para ver un ejemplo absolutamente detallado de la comprobación de transitividad, se recomienda ampliamente revisar el texto de Sushil (2017: 345).

Tabla 2. Comprobación de transitividad en la matriz de accesibilidad

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	1	1	1	1	1	1	1	1	1*	1	0
V2	0	1	1	1	1	1	1	1	1*	1	0
V3	0	0	1	1	1	1	1	1*	1	1	0
V4	0	0	0	1	1	1	1	1	1	1	0
V5	0	0	0	1	1	1	1	1*	1*	1	0
V6	0	0	0	0	0	1	1	1	1*	1*	0
V7	0	0	0	0	0	0	1	1	1	1	0
V8	0	0	0	0	0	0	0	1	0	0	0
V9	0	0	0	0	0	0	0	1	1	0	0
V10	0	0	0	0	0	0	0	1	0	1	0
V11	1	0	0	0	0	0	0	1	0	0	1

Fuente: López-Castro (2019)

Para cada nuevo enlace transitivo encontrado (1*), el cuestionario –base de conocimiento de lógica interpretativa– también se actualiza. Esto quiere decir que la entrada NO se cambiará a SÍ, y en la columna de interpretación se ingresa como transitivo. Para ilustrar este paso, el cuadro 5 muestra la actualización del cuestionario o base de conocimiento de lógica interpretativa (en su forma reducida, tal como se hizo en el cuadro 4), de acuerdo con los resultados de la matriz de accesibilidad transitiva (tabla 2). Observamos, por ejemplo, que relación V1 V9 ha resultado ser de enlace transitivo, por lo tanto, la base de conocimiento se actualiza sustituyendo NO por SÍ. Si hay una interpretación significativa disponible, se escribe junto con la entrada transitiva o, de lo contrario, se deja como está. Este paso se repite tantas veces como enlaces transitivos existan.

Cuadro 5. Base de conocimiento de lógica interpretativa (transitividad)

Número de comparaciones[a]	Número de variable	Comparación por pares de variables	SÍ/NO	¿Cómo o de qué manera la variable V1 mejora?
	V1	Las normativas existentes en su instituto tecnológico referidas a los contratos con las empresas MEJORAN…		
1	V1-V2	… la descentralización de su organización.	SÍ	Distribución de poder y responsabilidades, mecanismo de motivación para involucrados (libertad para la toma de decisiones), agilizar procesos
2	V1-V3	… la especialización que los miembros de su organización tienen en la actualidad.	SÍ	Nuevas responsabilidades, más especialización
3	V1-V4	… la comunicación entre el personal de su organización.	SÍ	Canales abiertos, informales y rápidos para responder a las necesidades de los interesados
4	V1-V5	… la confianza que tiene en su organización.	SÍ	Creación de normas *ad hoc* para incentivar y respaldar la colaboración con las empresas, el liderazgo directivo, la reputación local y nacional
5	V1-V6	… la colaboración de los miembros de su organización.	SÍ	Autoorganización, incentivos no monetarios para fomentar la colaboración interna
6	V1-V7	… los procedimientos para la colaboración con la industria en su instituto tecnológico.	SÍ	Reducir la burocracia en los procesos
7	V1-V8	… la complejidad y lentitud en los procesos.	SÍ	Atribución de responsabilidades al CIIT en procesos de vinculación (órgano descentralizado responsablede la gestión con las empresas)
8	V1-V9	… la falta de personal que apoye las tareas de vinculación.	No	Transitivo
			SÍ	Las normativas existentes de contratación son dependientes del nivel central
9	V1-V10	… la falta de profesionalización de los miembros de su organización.	SÍ	Nuevas responsabilidades, más especialización
10	V1-V11	… la rigidez de reglas impuestas por su institución o las agencias de financiamiento del gobierno.	NO	-
Número de comparaciones[a]	**Número de variable**	**Comparación por pares de variables**	**SÍ/NO**	**¿Cómo o de qué manera la variable V7 influye?**
	V7	**La falta de procedimientos para la colaboración con la industria en su instituto tecnológico, INFLUYE en...**		
71	V7-V1	… las normativas existentes en su instituto tecnológico referidas a los contratos con las empresas.	NO	-
72	V7-V2	… la descentralización de su organización.	NO	-
73	V7-V3	… la especialización que los miembros de su organización tienen en la actualidad.	NO	-
74	V7-V4	… la comunicación entre el personal de su organización.	NO	-
75	V7-V5	… la confianza que tiene en su organización.	NO	-
76	V7-V6	… la colaboración de los miembros de su organización.	NO	-
77	V7-V8	… la complejidad y lentitud en los procesos.	SÍ	Necesidad de búsqueda informal que ralentiza los procesos
78	V7-V9	… la falta de personal que apoye las tareas de vinculación.	SÍ	Carencia de directrices que permitan estructurar personal, presupuesto bajo y dependencia para contratar
79	V7-V10	… la falta de profesionalización de los miembros de su organización.	SÍ	Ausencia de guías y planes de formación a nivel sistema; a nivel interno se han ido creando
80	V7-V11	… la rigidez de reglas impuestas por su institución o las agencias de financiamiento del gobierno.	NO	-

Fuente: adaptado de López-Castro (2019)

En este paso, quien investiga verificará posibles inconsistencias en los enlaces transitivos. En caso de no detectar inconsistencias, se sigue con la aplicación de la técnica, pero, en caso contrario, se tendrá que verificar nuevamente cada par de variables de acuerdo con la regla (ver, Sushil, 2017: 345).

Paso 6. Partición de los niveles

Este paso se lleva a cabo para clasificar las variables en diferentes niveles y ver la importancia de las variables en el problema de investigación.

Los niveles se determinan de acuerdo con la ubicación de la mayor cantidad de 1 en la matriz de accesibilidad (recodificada para considerar la transitividad). Se empieza por identificar la(s) V_q que en su columna tengan la mayor cantidad de 1 consecutivos comenzando desde la parte superior, lo cual determina el primer nivel de jerarquía. Los niveles subsecuentes (que serán de menor jerarquía) se identifican con la misma regla o, en su caso, con las columnas qué tengan la misma distribución vertical de 1 y 0 (para este ejemplo, ver: Bamel, Dhir y Sushil, 2019). En consecuencia, todas las variables del sistema se agrupan en diferentes niveles.

El cuadro 6 muestra el ejemplo de la partición de niveles que se realiza a partir de la base la matriz de accesibilidad transitiva (tabla 2) para el caso de estudio (López-Castro, 2019).

Cuadro 6. Partición de niveles

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	1	1	1	1	1	1	1	1	1*	1	0
V2	0	1	1	1	1	1	1	1	1*	1	0
V3	0	0	1	1	1	1	1	1*	1	1	0
V4	0	0	0	1	1	1	1	1	1	1	0
V5	0	0	0	1	1	1	1	1*	1*	1	0
V6	0	0	0	0	0	1	1	1	1*	1*	0
V7	0	0	0	0	0	0	1	1	1	1	0
V8	0	0	0	0	0	0	0	1	0	0	0
V9	0	0	0	0	0	0	0	1	1	0	0
V10	0	0	0	0	0	0	0	1	0	1	0
V11	1	0	0	0	0	0	0	1	0	0	1
Niveles	**VII**	**VI**	**V**	**IV**		**III**	**II**	**I**	**II**		**VIII**

Número	Variable	Nivel
V8	Complejidad y lentitud en los procesos	I
V7	Falta de procedimientos establecidos para la colaboración con la industria	II
V9	Falta de personal que apoye las tareas	II
V10	Falta de profesionalización del personal universitario, especialmente los que gestionan las relaciones	II
V6	Colaboración intraorganizacional	III
V4	Comunicación intraorganizacional	IV
V5	Confianza intraorganizacional	IV
V3	Especialización de miembros para gestionar las actividades basadas en la experiencia	V
V2	Descentralización	VI
V1	Formalización existente	VII
V11	Rigidez de las reglas y regulaciones impuestas por las IES o agencias de financiamiento del gobierno	VIII

Fuente: adaptado de López-Castro (2019)

Paso 7. Dígrafo

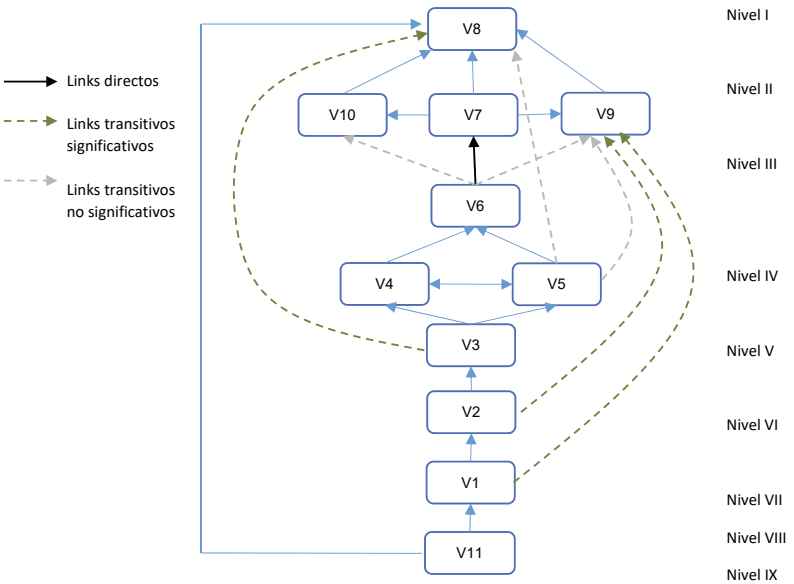
Este paso tiene como objetivo representar de manera visual las variables y sus interdependencias en términos de nodos y líneas. Se prepara sobre el cuadro 6 (partición de niveles) y el cuestionario –base de conocimiento de lógica interpretativa– (cuadro 5), con el fin de incluir los enlaces transitivos.

En este desarrollo, el factor de nivel superior se coloca en la parte superior del dígrafo y el factor de segundo nivel se coloca en la segunda posición y así sucesivamente, hasta que el nivel inferior se coloca en la posición más baja. Respecto de los enlaces transitivos, en la fase inicial se mostrarán todos en el dígrafo, para lograr así identificar qué enlaces transitivos pueden ser relacionados con una interpretación basada en la literatura o la evidencia sistematizada del campo: en el caso de que no se encontrara una relación interpretativa clara, el enlace transitivo

sería descartado. Para un correcto desarrollo del dígrafo, se recomienda revisar Sushil (2017: 345).

Un ejemplo de dígrafo inicial se muestra en el diagrama 2, construido a partir del caso seleccionado para ejemplificar. En su elaboración, los enlaces transitivos V1 V9; V2 V9 y V3 V8 se han mostrado porque existe una clara interpretación de ellos, pero los demás enlaces transitivos se identificaron como no relevantes, pues no se encontró que haya una interpretación significativa para el problema.

Diagrama 2. Dígrafo inicial



Fuente: adaptado de López-Castro (2019)

Paso 8. Matriz de interacción

El siguiente paso consiste en convertir el dígrafo inicial (en el que se han depurado los enlaces transitivos que carecían de interpretación clara) en una matriz de interacción (cuadro 7), con el objetivo de preparar el MEIT. La matriz de interacción incluye dos elementos:

a. Una matriz binaria, para indicar tanto las relaciones directas como los enlaces transitivos que son relevantes en el problema de investigación (cuadro 7(a)).

b. Una matriz interpretativa, para proporcionar la interpretación de las relaciones (cuadro 7(b)).

Por razones de espacio, la matriz interpretativa para el caso ilustrativo solo muestra las interpretaciones de las relaciones entre variables tratadas en pasos anteriores; sin embargo, cabe destacar que la matriz es extensa y, en su versión extensa, incluye todas las variables. El llenado de la matriz interpretativa se hará para todas las relaciones directas y transitivas significativas entre las variables, pues esta matriz es la que hará que la elaboración del MEIT sea más sencilla y, a su vez, ayudará a quien investiga a proporcionar una explicación completa y detallada de las relaciones del sistema de variables del problema.

Cuadro 7. Matriz de interacción

(a) Matriz binaria

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	-	1	1	1	1	1	1	1	1*	1	0
V2	0	-	1	1	1	1	1	1	1*	1	0
V3	0	0	-	1	1	1	1	1*	1	1	0
V4	0	0	0	-	1	1	1	1	1	1	0
V5	0	0	0	1	-	1	1	0	0	1	0
V6	0	0	0	0	0	-	1	1	1	0	0
V7	0	0	0	0	0	0	-	1	1	1	0
V8	0	0	0	0	0	0	0	-	0	0	0
V9	0	0	0	0	0	0	0	1	-	1	0
V10	0	0	0	0	0	0	0	1	1	-	0
V11	1	0	0	0	0	0	0	1	0	0	-

(b) Matriz interpretativa

	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11
V1	-	Distribución de poder y responsabilidades, mecanismo de motivación para involucrados (libertad para la toma de decisiones), agilizar procesos	Nuevas responsabilidades, más especialización	Canales abiertos, informales y rápidos para responder a las necesidades de interesados	Creación de normas *ad hoc* para incentivar y respaldar la colaboración con empresas, liderazgo directivo, reputación local y nacional	Autoorganización, incentivos no monetarios para fomentar la colaboración interna	Reducir la burocracia en los procesos	Atribución de responsabilidades al CIIT en procesos de vinculación	La contratación es dependiente	Nuevas responsabilidades, más especialización	
V2		-									
V3			-								
V4				-							
V5					-						
V6						-					
V7							-	Necesidad de búsqueda informal que ralentiza los procesos	Carencia de directrices; presupuesto bajo y dependencia para contratar	Ausencia de guías, planes de formación a nivel sistema; a nivel interno se han ido creando	
V8								-			
V9									-		
V10										-	
V11											-

Notas: En 7(a), las negritas señalan los enlaces directos; las cursivas, los enlaces transitivos significativos, y el subrayado, el enlace transitivo marcado como 1 que cambia a 0 debido a que la interpretación no es significativa en el problema.

Fuente: adaptado de López-Castro (2019)

Paso 9. Desarrollo de un modelo estructural interpretativo total

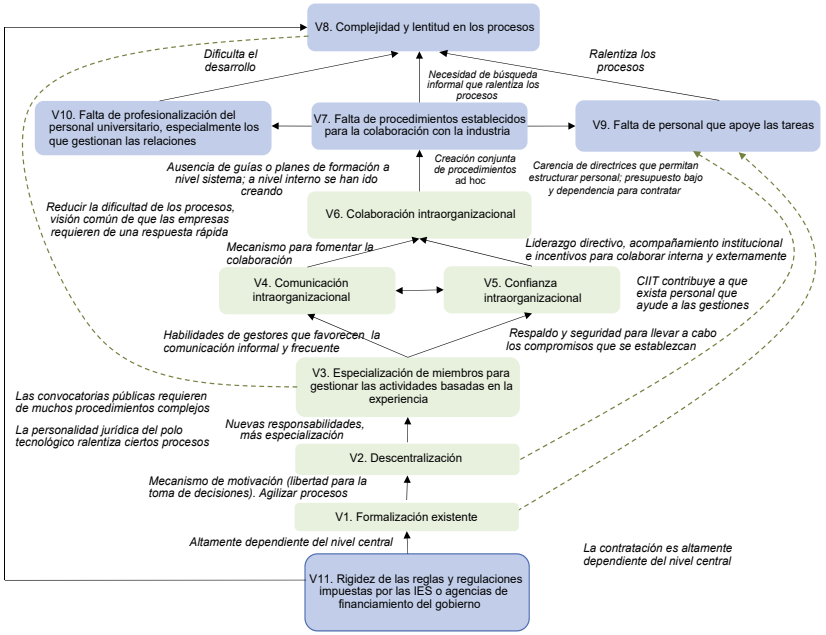
En este último paso se desarrolla el MEIT con el objetivo de ofrecer una imagen muy clara del sistema de variables o problema de investigación, y su flujo de relaciones. Para su elaboración, se toma el dígrafo inicial depurado (extraído del diagrama 2) y las interpretaciones de la matriz de interacción (cuadro 7(b)).

Los nodos en el dígrafo, que hasta ahora estaban indicados con las abreviaciones de las variables, se reemplazan con la descripción de la variable real y, en cada relación direccional, se proporciona la interpretación correspondiente. El diagrama 3 muestra el MEIT para el caso ilustrativo; es aquí que quien investiga puede tener una comprensión completa del problema de investigación planteado y está en condiciones de responder el "qué", "cómo" y "por qué" del problema de investigación planteado.

López-Castro encontró que los elementos definidos como parte de la estructura organizacional del Tecnológico Nacional de México en Celaya (señalados en color verde), tienen un efecto en la reducción de los obstáculos en el proceso de vinculación y transferencia de conocimiento con las empresas (señalados en color azul). También se encontró la existencia de un obstáculo completamente exógeno al tecnológico (V11: rigidez de las reglas y regulaciones impuestas por las IES o agencias de financiamiento del gobierno) que tiene influencia en la gestión y operativa de los procesos internos.

La validación última del MEIT se realiza en el contexto en el que se inserta el problema o unidad de análisis, es decir, a través de la implementación. Se les pide a los y las participantes de campo que verifiquen el modelo, si es necesario, se realizarán ajustes en las interpretaciones, pues tanto los niveles como las relaciones identificadas ya han sido verificadas en las fases anteriores del método. Finalmente, en la elaboración del MEIT hay algunos errores que pueden surgir en el modelado. Para verificar que el modelo esté desarrollado de la manera correcta, se recomienda ampliamente revisar Sushil (2018: 482).

Diagrama 3. Modelado estructural interpretativo total



Fuente: López-Castro (2019)

Análisis estructural-causal cualitativo (AECC)

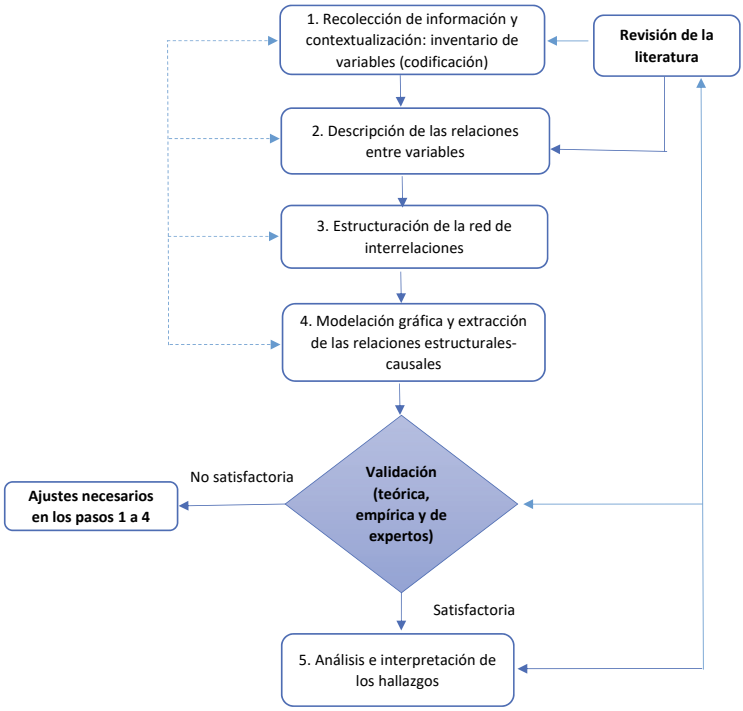
A grandes rasgos, el AECC es un método que permite transitar del problema de investigación a la entidad relacional detrás de este. Más aún, este análisis permite capturar y organizar la evidencia empírica –observada en la realidad–en un objeto de estudio construido a partir de la contextualización, categorización y conceptualización de dicha información. Se parte de la premisa de la existencia de una estructura (formada por relaciones causales) intrínseca y poco evidente en los fenómenos sociales (von Wright, 1987; Maxwell, 2004; Pearl, 2009). Utiliza un proceso inductivo-deductivo (*ida y vuelta*) para la abstracción de las relaciones teóricas entre las variables o factores de interés del fenómeno en estudio.

Blalock (1961) menciona cuatro criterios fundamentales para establecer una relación causal: la universalidad (todo efecto tiene su causa), la necesariedad (dadas las condiciones necesarias y suficientes para que se produzca un efecto, este necesariamente se produce, implicando la contextualización del problema de investigación; la univocidad (a

determinadas causas, determinados efectos) y la temporalidad (la causa siempre antecede al efecto). Criterios a los que Judea Pearl (2009) añade la especificidad de la causa (para un solo efecto se plantea una sola causa) y la plausibilidad (la relación entre causa y efecto es concordante con la teoría). El cumplimiento de estos criterios permitiría establecer, en investigaciones cualitativas, relaciones estructurales entre los factores o variables de interés y, más aún, determinar el orden causal que siguen estas relaciones (von Wright, 1987; Pearl, 2009).

Siguiendo el planteamiento de Curnow, McLean y Shepherd (1976), el diagrama 4 presenta un diagrama de las etapas que se proponen seguir para la implementación del AECC cualitativo.

Diagrama 4. Etapas del AECC para investigaciones cualitativas



Fuente: elaboración propia con base en Curnow, McLean y Shepherd (1976), Pearl (2009) y Martínez (2014)

Descripción del proceso del AECC y su aplicación en un ejemplo ilustrativo

El AECC comprende cinco etapas, más un proceso de validación y ajuste. A continuación, se describe cada una de estas etapas, mediante su aplicación en el caso de una innovación en telemedicina en México (para un análisis completo del caso ver Martínez *et al.* 2018).

Recolección de información y contextualización: inventario de variables (codificación)

En esta primera etapa se busca recabar la mayor información posible del fenómeno en estudio, contextualizándolo en un tiempo y espacio. Consiste en definir el sistema a analizar y sus elementos pertinentes; para ello, se deberá acotar la realidad social, delimitarla conceptual, espacial y temporalmente. Se trata de estudiar una estructura social concreta, de una zona, de un país, de una región, etc., en un momento determinado, sin obviar los condicionantes históricos.

Al delimitar los conceptos clave, se construyen variables, indicadores y/o categorías analíticas que sirven para la descripción y comprensión de dicha realidad social. Entonces se procede a realizar un inventario de todas las variables o categorías que caracterizan el problema particular de estudio. Es conveniente ser lo más exhaustivo posible en esta primera etapa, teniendo cuidado de no dejar nada sin explicar al describir el problema de estudio.

En esta etapa se organiza y codifica la información existente con base en la teoría y la propia evidencia empírica que aportan nuevas categorías de análisis (Danermark *et al*., 2002; Barco y Carrasco, 2018). Se extraen las variables o elementos (hechos, actores y acciones) relevantes de acuerdo con el objetivo de estudio. Posteriormente, estos elementos se clasifican y/o codifican de acuerdo con la teoría, los trabajos empíricos anteriores y la evidencia,[15] para luego operacionalizar y agrupar las múltiples variables o elementos en un conjunto de factores más reducido (Pearl, 2009).

Aunque los AECC permiten deducir nuevas categorías analíticas y relaciones causales, toda "nueva teoría" requiere de la base de los hallazgos y proposiciones de estudios anteriores, para llevar a cabo una contrastación o adición a la teoría existente (Pearl, 2009; Blersch *et al*. 2021). Así, el AECC va de la teoría a la realidad estudiada, pero también de la evidencia a la teoría, en un proceso de *ida y vuelta.*

15 Para una explicación detallada del proceso de codificación, ver: Strauss (1987) y Maxwell (1996).

El cuadro 8 es un pequeño extracto de la operacionalización de las categorías analíticas del caso ilustrativo y, además, resume el proceso de codificación de las variables o categorías seleccionadas. Las dos primeras columnas muestran las dimensiones de análisis y sus categorías analíticas correspondientes (que representan las variables observadas). En la tercera columna se visualizan los ítems mediante los cuales se "observan" dichas categorías.

Cuadro 8. Categorías analíticas del caso de telemedicina en México

Dimensión	Categoría analítica (variables observadas)		Operacionalización en ítems/pregunta
	General	Específica	
Aprendizaje interactivo	Innovación orientada por la escasez	Identificación de necesidades sociales	¿Con base en qué criterios/motivos se elige la población/comunidades (beneficiarios) a la que se quería llegar con el proyecto?
			¿Hubo algún tipo de intervención de otro actor para decidir las características y lineamientos (a quién iba dirigida y por qué) del proyecto, además de Médica Sur y Telemed?
		Dinamización de la demanda de conocimiento	¿Por qué razones se decide realizar un proyecto de telemedicina?
			¿Cuál fue el objetivo del proyecto?
			¿Se buscaba explícitamente lograr una innovación?
	Innovación como resultado de la interacción	Incentivos de política pública a la innovación	¿Quién financió el proyecto? ¿El Estado tuvo alguna participación en ello?
		Incentivos de política pública a la vinculación	¿El programa que financió el proyecto promovía de algún modo la vinculación con otro tipo de actores? Si así fue, ¿de qué forma?
		Oportunidad de negocio	¿El proyecto, en sí mismo, era visto como un negocio?
			Sin el financiamiento recibido, ¿el proyecto de todas formas se habría realizado?
		Interacciones sistémicas y coordinadas	¿Por qué se buscó la vinculación con esos actores (razones)?
			¿Con qué instituciones se estableció contacto?
Innovación inclusiva	Transferencia de tecnología	Transferencia y adopción tecnológica de los usuarios	¿Cómo fue el proceso de implementación tecnológica en los hospitales (usuarios)? Capacitación, programas de educación a la población usuaria, etc.
	Inclusión social	Mejoramiento en el acceso a bienes y servicios de las poblaciones excluidas	¿Realmente se llegó a la población objetivo (cifras)?
			¿Se considera que esta innovación tuvo un impacto en el mejoramiento en el acceso a servicios de salud de las comunidades menos favorecidas?

Fuente: adaptado de Martínez *et al*. (2018)

Para lograr recabar la mayor información sobre el problema de investigación, sus actores, roles, interacciones, hechos clave, etc. existen diversas herramientas de recolección propias de la investigación cualitativa.

Algunas de las herramientas más utilizadas son: entrevistas semiestructuradas a informantes calificados, observación directa o participante, documentación, etnografía y narrativas (Pearl, 2009). Asimismo, es posible que el AECC se apoye en los resultados de la aplicación previa de otro método de análisis, como puede ser un estudio de caso, pero esto no es una condición necesaria sino una elección de quien investiga de acuerdo con los objetivos de su trabajo.

En el ejemplo, el objetivo del estudio fue doble. Por un lado, buscó analizar el proceso de una innovación de productos inclusivos (software y equipos de telemedicina), identificando a los actores y hechos determinantes en este tipo de innovaciones. Por otro lado, extraer y estilizar las relaciones causales entre los determinantes de dicho proceso. De acuerdo con estos objetivos, se optó por una metodología fenomenológica que combinó el estudio de caso (Yin, 2003) con el AECC (Burt, 1982; Pearl, 2009; Maxwell, 2021).

Debido a que el caso a analizar era un hecho pasado, este fue reconstruido y se dividió en cinco etapas cronológicas. A través de estas etapas se fueron identificando los actores clave y su rol en el caso de estudio, sus interacciones, así como los hechos y acciones determinantes en el proceso de desarrollo de la innovación en telemedicina.

Las herramientas de recolección de información fueron: 1) trece entrevistas semiestructuradas a informantes calificados; 2) documentos oficiales del proyecto (reporte técnicos y financieros) y 3) otras fuentes secundarias externas. Las entrevistas a informantes calificados fueron la principal fuente de información; de hecho, como se aprecia en la cuarta columna del cuadro 8, la unidad mediante la cual se observó cada categoría analítica propuesta fueron ciertas preguntas construidas en función de la categoría que se quería observar o hacer tangible.

Descripción de las relaciones entre variables

En una segunda etapa, el método estructural busca la explicación del sistema (sus características y su dinámica) en una configuración subyacente de elementos que permite su interpretación deductiva (Burt, 1982). El descubrimiento de la estructura es el segundo momento del análisis estructural. Se trata pues de "hacer hablar" a la estructura, es decir, de extraer de ella las leyes relacionales que definen el sistema como totalidad estructurada, como actividad estructurante y como proceso cambiante; así como analizar y determinar las relaciones existentes entre los

elementos del sistema, o sea, entre las categorías analíticas, en el caso de telemedicina estudiado.

Una de las formas más utilizadas en esta etapa es la construcción de una matriz de análisis estructural (cuadro 9), la cual consiste en vincular las variables en una tabla de doble entrada, preparada especialmente para el caso en estudio. Las filas y columnas en esta matriz corresponderán a las variables o categorías codificadas en la primera etapa.

Previo a la elaboración de la matriz estructural, se ubicaron y clasificaron a los actores clave del caso de telemedicina con el fin de describir el rol y la función de cada actor (ver cuadro 9). Es imperante mencionar que las funciones desempeñadas por los actores y sus relaciones son las que dan lugar a las categorías analíticas, es decir, son el insumo de la matriz estructural. El cuadro 9 muestra un extracto de la sistematización de actores y sus roles o funciones en el caso ilustrativo.

Cuadro 9. Los actores del caso: funciones y roles jugados

Sectores de la sociedad		**Actores**		**Función/rol en el desarrollo de la Telemedicina y en la RIIT**
A	**Gobierno**	a1	Gobierno Federal y Secretaría de Salud	Dinamizador de la demanda de conocimiento
		a2	CONACyT	Orientador y facilitador de recursos financieros.
B	**Productores de conocimiento científico y tecnológico**	b1	Universidad La Salle	Nodo de vinculación
		b2	Universidad Panamericana	Nodo de vinculación
		b3	CICESE	Nodo de la Red como aliado tecnológico
		b4	CUDI	Aliado para la formación de infraestructura de telecomunicaciones
		b5	UAM-C	Desarrollador del modelo de transferencia tecnológica implementado
C	**Sector productivo (empresas)**	c1	Médica Sur	Nodo supervisor del proyecto
				Desarrollador de conocimiento endógeno
		c2	Telemed	Líder y administrador del proyecto
		c3	Telmex	Proveedor de servicios de internet

	Demanda	**Actores vinculados con el problema de exclusion en el acceso al servicio de salud en México**		
D	**Usuarios**	d1	Clínica MAS de Tlapa, Guerrero	Nodo del programa de cirugía teleasistida y usuario de la tecnología
		d2	HRAE de Oaxaca	Nodo del programa de cirugía teleasistida y usuario de la tecnología
		d3	HRAE del Bajío	Nodo del programa de cirugía teleasistida y usuario de la tecnología
		d4	Hospital Básico Comunitario de Larráinzar, Chiapas	Nodo del programa de cirugía teleasistida y usuario de la tecnología
		d5	Hospital GEA González	Nodo central y usuario de la tecnología
				Nodo de vinculación
		d6	Hospital General de Milpa Alta	Usuario de la tecnología
	Beneficiarios	d7	Pacientes de los hospitales	Población objetivo (beneficiarios)

Nota: Las letras con las que se clasifica a cada actor tienen una doble función: 1) clasificar a cada actor con su correspondiente macroactor y 2) porque posteriormente, en el siguiente apartado, se muestra el diagrama causal que se sustenta utilizando esta notación de los actores.

Fuente: tomado y adaptado de Martínez *et al.* (2018)

Una vez sistematizadas las funciones de cada actor, se extraen los hechos clave (variables o categorías analíticas) que, en última instancia, son los determinantes de una innovación inclusiva de producto. Estas categorías son el insumo de la matriz estructural (cuadro 10) mediante la cual se hace posible relacionar las categorías analíticas (conceptos: acciones y hechos) que se presentan en el cuadro 8, logrando visualizar y hacer explícitas las relaciones directas e indirectas entre estas, para posteriormente esquematizarlas en forma estilizada y gráfica en la siguiente etapa.

Estructuración de la red de interrelaciones

Ya explícitas las relaciones entre las categorías o variables del fenómeno social en estudio, se procede, en una tercera etapa, a describir el tipo (directas, indirectas, bidireccionales o espuria) y grado de estas relaciones; así como a organizarlas en una estructura lógica con base a la teoría revisada y a la evidencia empírica recabada.

En el caso ilustrativo, la matriz estructural muestra la interacción de las 18 variables identificadas en el caso de estudio, arrojando un total de

324 relaciones posibles, que resultan de relacionar cada factor con cada uno de los otros factores (18 x 18), 24 fueron relaciones directas (representadas con el número 1 y el signo *).

En resumen, para el caso de una innovación en telemedicina, la matriz estructural mostró que el proceso de generación y desarrollo de innovaciones en productos y servicios inclusivos no es lineal ni unicausal, ya que de otro modo los números 1 formarían una diagonal perfecta (cada acción sería causada únicamente por la acción inmediatamente anterior); sino que se presenta una causalidad múltiple que pone en evidencia la complejidad detrás del proceso de generación de bienes y servicios inclusivos que surge a través de las interacciones entre múltiples actores diversos.

Cuadro 10. Matriz de análisis estructural

	Categorías	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI	XVII	XVIII
I	Identificación de necesidades sociales		1	1															
II	Dinamización de la demanda de conocimiento			*					1										
III	Incentivos de política pública a la innovación		*		1	1													
IV	Incentivos de política de CTI a la vinculación						1												
V	Dinamización de la oferta de conocimiento						1				1								
VI	Vinculaciones/interacciones							1		1									
VII	Inversión e investigación compartida									*	1	1							
VIII	Oportunidades de negocio					1													
IX	Articulación de capacidades							*				1							
X	Tecnología externa											1							
XI	Generación de conocimiento endógeno												1						
XII	Innovación local/endógena													1					
XIII	Registro de derechos de propiedad														*				
XIV	Transferencia tecnológica (a usuarios)													*		1			
XV	Adopción tecnológica (de usuarios y beneficiarios)																1		
XVI	Mejoramiento en el acceso a ByS																	1	1
XVII	Reducción de privaciones instrumentales																		
XVIII	Generación de oportunidades																		

Notas: (1) Relación directa; (*) relación bidireccional. Debido a que las relaciones espurias fueron descartadas, las celdas vacías representan relaciones indirectas.

Fuente: recuperado de Martínez *et al*. (2018)

La lectura de la matriz estructural se realiza en un espacio de doble entrada. El cruce o intersección de dos variables se representa en la matriz estructural como una celda (intersección de columnas y filas). Es así que, toda vez que las filas y columnas de dicha matriz representan una categoría analítica o variable, esta herramienta nos va mostrando la interacción de cada variable con todas las demás (incluyéndose ella misma), es decir, se van deduciendo las relaciones estructurales subyacentes al fenómeno complejo estudiado.

Sin embargo, la matriz *per se* no nos da un orden causal de las relaciones encontradas. Para establecer relaciones de causa-efecto es necesario hacer uso de la evidencia y dar una secuencia a los hechos; es por esto que la recolección y contrastación de la evidencia con diversas fuentes de información, deberá ser minuciosa. Por lo tanto, la esquematización gráfica ayuda a estructurar dichas relaciones, siempre tomando en cuenta los criterios para establecer una relación causal (Blalock, 1961 y Pearl, 2009), en la que la temporalidad de los hechos y la contrastación teórica serán clave para lograr ordenar las variables y sus relaciones, es decir, para lograr una modelación cualitativa válida y plausible.

Modelación gráfica, extracción de relaciones causales y proceso de validación

En esta cuarta etapa se procede a pasar de la matriz de análisis estructural a una forma gráfica, en la que se esquematizan las relaciones causales detrás del fenómeno social analizado, es decir, que se hace evidente y visible su estructura interna. Así, al igual que ocurre en el MEIT, el AECC recurre a la teoría de grafos para la estilización de los resultados. De acuerdo con Pearl (2009) y Maxwell (2021), dicha teoría proporciona una herramienta de notación matemática para los conceptos y sus relaciones que no se expresan fácilmente en ecuaciones algebraicas.

Un grafo o diagrama (G = (V, E)) se define por medio de un conjunto V finito de vértices o nodos y un conjunto E (v x v) de aristas que conectan los vértices. Los vértices representarán las variables o categorías elegidas y las aristas o flechas indicarán relaciones entre los vértices, describiendo la lógica estructural-causal del sistema o fenómeno estudiado. El gráfico debe reproducirse de manera tal que contribuya a desenmarañar con rapidez la red de interrelaciones dentro del sistema, es decir, debe transmitir más cosas que su matriz original (Curnow, McLean y Shepherd, 1976 y Pearl, 1995).

En el AECC, este grafo es denominado como *diagrama causal*, en el que se estilizan gráficamente las relaciones estructurales extraídas, las cuales se categorizan en: a) directas (e1 causa e2), b) indirectas (e1 causa e3 que causa e2), c) bidireccionales (e1 causa e2 y e2 causa e1) y d) espurias (e3 causa e1 y e2) (Pearl, 1995).

Así, para el caso de telemedicina, el diagrama 5 presenta el diagrama causal que estiliza gráficamente el orden causal que siguen estas relaciones estructurales encontradas. Mediante este diagrama se logra una generalización analítica de las acciones, hechos y relaciones que llevaron a mejorar el acceso a ciertos bienes y servicios de calidad, a través de la generación y aplicación creativa y efectiva del conocimiento (innovación).

Adicionalmente, el diagrama 5 también permite visualizar y ubicar la participación de cada actor en dicho proceso, y observar cómo la acción de un actor determina o habilita las acciones de otros actores, dando lugar a las relaciones estructurales entre los sectores sociales y/o actores involucrados (ver el cuadro 9). Así, estas interrelaciones van formando una cierta estructura sistémica que evoluciona y coevoluciona en un contexto de retroalimentación y aprendizaje interactivo.

Finalmente, un paso de suma importancia es garantizar la validez teórica y de constructo del modelo propuesto. Debido al tipo de información (cualitativa), la validación en el AECC se realiza mediante la contrastación teórica y la opinión de expertos en el tema.

Como se ha mencionado a lo largo de este trabajo, los modelos estructurales sustentan sus afirmaciones causales y la construcción de sus categorías analíticas en la teoría existente o en hallazgos de estudios empíricos pasados. Así, la validación teórica de este análisis consistió en poner a prueba las categorías y las relaciones causales encontradas, mediante su contrastación con la teoría revisada. Ahora bien, la validez del constructo supone establecer medidas operacionales correctas para los conceptos a ser estudiados, se trata de definir con precisión qué se quiere observar (categorías o variables), a la vez que deben establecerse las razones por las cuales las fuentes de información que se utilizan son las mejores para ello. Como se mencionó anteriormente, en el caso ilustrativo se utilizaron múltiples fuentes de información (internas y externas), cuya evidencia fue debidamente triangulada para su validación empírica.

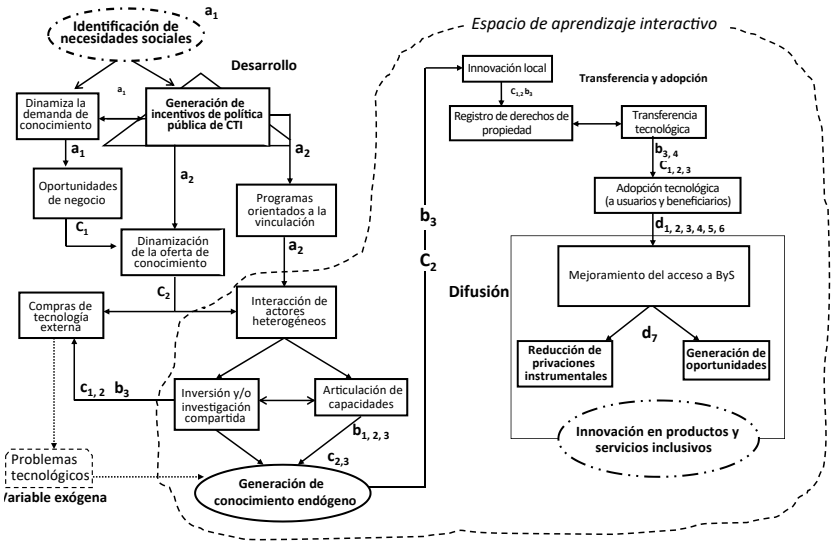
A efectos de garantizar la validez del constructo, se establecieron cadenas de evidencia, con el fin de triangular la información obtenida de las diversas fuentes de información (Yin, 2003), lo cual implica que los diferentes hechos del caso fueron sustentados y corroborados por más de

una fuente de información. Para este trabajo ello significa ordenar cronológicamente las acciones de los actores para observar si existe algún patrón en el desempeño de estos y en las actividades que ellos desarrollan: ¿la acción realizada por el actor B fue influenciada de alguna manera por la acción del actor A? Así, es posible establecer un cierto orden causal detrás de las relaciones entre los actores.

Más allá de que cada una de las variables o categorías fueron definidas conceptualmente y a la luz de la teoría, y que cada una de ellas y sus relaciones fueron respaldadas también por la evidencia; es importante la revisión de los resultados obtenidos a través de un panel de expertos (actores clave del fenómeno social analizado, jurado evaluador, comité dictaminador, etc.) para aumentar su validez y confiabilidad.

Para la evaluación de la validez de las relaciones causales que se desprendan del análisis, es central que haya un cumplimiento riguroso de los criterios de especificidad, necesariedad, temporalidad y plausibilidad, ya que de ello dependerá la generalización teórica y analítica de los resultados obtenidos.

Diagrama 5. Diagrama causal: relaciones estructurales entre los determinantes de la innovación inclusiva



Fuente: recuperado de Martínez *et al.* (2018)

Análisis e interpretación

Como su nombre lo dice, en esta última etapa se procede al análisis e interpretación de los hallazgos/resultados a la luz de la teoría, de la evidencia empírica y del contexto específico del problema de estudio. Se trata de discutir los resultados con la literatura revisada, ver cuál es el aporte de los hallazgos y cuáles sus implicaciones. Vale señalar nuevamente que la generalización de los resultados de este tipo de metodología es analítica y contextual/local. Sin embargo, las categorías propuestas (y sus relaciones), al ser constructos teóricos, pueden ser aplicables a estudios de un mismo fenómeno social complejo.

Discusión: diferencias, ventajas y limitaciones de las metodologías estructurales propuestas

Los MEC ayudan a modelar o representar un sistema complejo de manera simplificada, facilitando la identificación de la estructura dentro de un sistema (Gardas, Narkhede y Raut, 2017). En otras palabras, son herramientas para hacer de una investigación de corte cualitativo una fuente de información para la toma de decisiones al permitir explicaciones contextualizadas y específicas del problema de estudio.

Aunque el objetivo de ambas propuestas es el mismo, presentan diferencias metodológicas significativas (ver cuadro 11). Se observa, por ejemplo, que el MEIT ofrece la posibilidad de establecer distintos tipos de relaciones entre variables (ver cuadro 2) mientras que el AECC se centra en el estudio de relaciones de causa y efecto. Otras de las diferencias es que el MEIT se centra en el análisis de problemas muy específicos utilizando variables homogéneas al problema de interés (unidimensional), mientras que el análisis estructural-causal cualitativo se centra en el análisis de realidades desconocidas incluyendo variables y actores (multidimensional). Finalmente, otra de las diferencias relevantes radica en que la aplicación del MEIT se realiza a través de pasos estructurados y concretos para descubrir y modelar relaciones entre variables, mientras que el AECC también lo hace, pero de una manera más flexible.

Cuadro 11. Diferencias entre el MEIT y el análisis estructural-causal cualitativo

	MEIT	AECC
Tipo de pregunta que responde	¿Cómo? y ¿por qué?	¿Cómo? y ¿por qué?
Tipo de relación entre variables	Varios tipos de relaciones predeterminadas	Causa-efecto
Criterios causales	No	Sí
Análisis de estructuras	Problema concreto a través de variables/elementos homogéneos (unidimensional)	Descubre fenómenos desconocidos, incluyendo diversas dimensiones analíticas (multidimensional) como: actores, funciones y categorías/variables y sus relaciones causales
Etapas del proceso	9	5
Herramienta de recolección de información	Cuestionario con preguntas predeterminadas; entrevistas semiestructuradas, focus group	Cualquier herramienta o método de corte cualitativo
Validación	Grupos de expertos	Teoría, evidencia empírica, grupo de expertos

Fuente: elaboración propia

Más allá de las ventajas comentadas durante este trabajo, es necesario reconocer que al ser métodos emergentes en las ciencias sociales y al haber aún pocos estudios que los apliquen, ambos presentan una serie de limitaciones asociadas a su aplicabilidad y comprobación de validez. El valor añadido de utilizar estos modelos interpretativos es la posibilidad de clarificar las relaciones estructurales difusas, en muchos casos, del análisis cualitativo (Attri, Dev y Sharma, 2013: 3). Esta metodología ha sido criticada bajo el argumento de que la relación entre las variables siempre depende del conocimiento de la persona que las juzga y no tiene validez estadística. En este sentido, algunos investigadores argumentan que la limitación de no ser estadísticamente válida es discutible, ya que su creciente aplicación en diversas áreas de investigación la establece como una "herramienta analítica de apoyo que es capaz de desarrollar un modelo inicial" (Menon y Suresh, 2020: 261). Más aún, Pearl (2009) menciona que estos métodos contribuyen a incrementar el conocimiento, al permitir la construcción de nuevas categorías analíticas basadas en un enfoque inductivo-deductivo.

Adicionalmente, se han propuesto diversas alternativas para superar el problema de la validez estadística. El MEIT puede utilizarse de forma articulada con el modelado de ecuaciones estructurales (MEE) (en inglés, Structural Equation Modelling (SEM)) para validar estadísticamente los

modelos teóricos propuestos a partir del MEIT.[16] Podemos decir entonces que el MEIT y el MEE pueden ser complementarios entre sí, lo cual es una ventaja de utilizar este método. Es importante comentar que el uso de esta técnica plantea distintos desafíos para modelar problemas. El primero se refiere al número de variables involucradas en el problema, es decir, cuánto más variables identificadas, mayores serán el número de comparaciones y, por lo tanto, más complejidad en la aplicación de la técnica. Otro desafío para aplicar la técnica en un contexto específico es determinar quiénes serán los y las participantes, porque de su familiaridad con el problema y conocimiento depende la validez de los resultados. Finalmente, la corrección de estos modelos debe realizarse de forma muy cuidadosa: se ha observado que los MEIT pueden tener diferentes tipos de errores o inconsistencias, sobre todo, las relacionadas con la partición de niveles y la regla de transitividad; quien investiga deberá prestar atención a estas dos cuestiones para evitar una representación incorrecta del modelado del problema. Aunque en este capítulo hemos proporcionado pautas claves para evitar estos errores, se sugiere a quienes tengan interés en utilizar la técnica consultar los trabajos de Sushil (2017, 2018), quien proporciona una guía imprescindible para verificar la corrección en cada etapa del modelado del MEIT y representar correctamente los modelos.

En el caso del AECC, al ser un método fundamentado en la contextualización, la teoría y la abstracción de la realidad, sus resultados y hallazgos son específicos al estudio (casos, hechos, actores, etc.). Por lo tanto, si se quiere replicar un estudio basado en dicho análisis, la metodología seguida deberá ser debidamente adaptada al contexto.

Conclusiones

Este trabajo argumenta que los modelos estructurales cualitativos tienen un gran potencial de aplicación en la investigación social, en la que existe más información (empírica) que variables disponibles, tanto teóricamente como a nivel de datos o mediciones cuantitativas. Más aún, se afirma que, aunque con las limitaciones inherentes a cualquier método,

16 Otra alternativa que se propone utilizar, es la técnica MICMAC, también conocida como Matrice d'Impacts Croisés Multiplication Appliqués à un Classement (por ejemplo: Ali y Dubey, 2014; Gandhi, 2015) para clasificar las variables en función del poder que tienen en el problema.

los MEC pueden ser una metodología robusta para el ordenamiento y sistematización de información compleja, dispersa, cualitativa y difícilmente parametrizable, así como para su análisis y para la obtención de resultados válidos, confiables y analíticamente generalizables.

Particularmente, el estudio y la comprensión de los procesos de CTI son una tarea compleja por el gran número de elementos que están interrelacionados y por el tipo de información cualitativa que implican estos procesos. Esta complejidad es difícilmente parametrizable con datos cuantitativos y, en este contexto, las aproximaciones cualitativas son necesarias. Uno de los beneficios del uso de investigaciones de corte cualitativo de manera general es que permite comprender procesos sociales complejos. Además, permite capturar aspectos esenciales de un fenómeno desde la perspectiva de los participantes del estudio, descubrir creencias, valores y motivaciones que subyacen a los comportamientos de un problema determinado. En otras palabras, la investigación cualitativa busca generar conocimientos novedosos y describir la complejidad, amplitud o variedad de ocurrencias o fenómenos, pero también dar explicaciones a problemas y/o procesos localmente determinados. Es así que, una comprensión clara de la metodología cualitativa y la incorporación sistemática de técnicas establecidas para garantizar el rigor, como los MEC, provee a quienes investigan en las ciencias sociales de herramientas poderosas y novedosas para analizar, comprender y explicar los procesos de CTI.

En esa dirección, este capítulo ha ofrecido dos alternativas metodológicas: el modelado estructural interpretativo total (MEIT) y el análisis estructural-causal cualitativo (AECC). Ambos métodos están diseñados para interpretar estructuras, dimensionalidad y relaciones subyacentes en los fenómenos complejos; utilizan datos cualitativos que sirven para operacionalizar conceptos, arrojando luz sobre fenómenos y problemas multidimensionales y multicausales. Sin embargo, la diferencia más significativa entre ambos métodos es que el MEIT, desde el enfoque interpretativo de la causalidad, permite descubrir y modelar diversos tipos de relaciones entre variables a través de un proceso estructurado, por ejemplo: conocer la causa, los problemas, las oportunidades, etc., en un problema determinado y su análisis es unidimensional, es decir, analiza variables o dimensiones de interés de naturaleza homogénea, siendo muy útil para la descripción de un proceso de forma clara y precisa. El análisis estructural-causal por su parte, desde el pospositivismo, se centra en descubrir y modelar relaciones de causa y efecto a través de un

proceso más flexible y su alcance es multidimensional, porque es capaz de identificar relaciones entre variables y actores (naturaleza heterogénea) descubriendo la causalidad enmarcada en un espacio y tiempos determinados (contexto).

Finalmente, este tipo de modelos tiene un gran potencial para el análisis y explicación de los problemas específicamente localizados. Como se mencionó a lo largo de este capítulo, los MEC operan con evidencia siempre contextualizada en un tiempo y espacio determinado, por lo tanto, la replicación de estos estudios deberá tener en cuenta los elementos específicos al problema (tiempo, espacio, y otros elementos particulares como los actores y hechos relevantes).

Bibliografía

Akriti, J.; Sharma, R. y Vigneswara, I. (2018). "Total Interpretive Structural Modelling of Innovation Measurement for Indian Universities and Higher Academic Technical Institutions". En Kulkarni, A. J.; Singh, T. P. y Sushil (eds.), *Flexibility in Resource Management*, pp. 29-53. Singapore: Springer.

Ali, S. y Dubey, R. (2014). "Identification of flexible manufacturing system dimensions and their interrelationship using total interpretive structural modelling and fuzzy MICMAC analysis". *Global Journal of Flexible Systems Management*, vol. 15, nº 2, pp. 131-143.

Attri, R.; Dev, N. y Sharma, V. (2013). "Interpretive structural modelling (MEI) approach: An overview". *Research Journal of Management Sciences*, vol. 2, nº 2, pp. 3-8.

Bamel, N.; Dhir, S. y Sushil, S. (2019). "Inter-partner dynamics and joint venture competitiveness: a fuzzy MEIT approach". *Benchmarking An International Journal*, vol. 26, nº 2.

Barco, B. y Carrasco, A. (2018). "Explicaciones causales en la investigación cualitativa: elección escolar en Chile". Magis. *Revista Internacional de Investigación en Educación*, vol. 11, nº 22, pp. 113-124.

Blalock, H. M. (1961). "Correlation and causality: the multivariate case". *Social Forces*, vol. 139, pp. 246-251.

Blersch, R.; Bonnell, T.; Franchuk, N.; Lucas, M.; Nord, C. y Varsanyi, S. (2022). "Causal analysis as a bridge between qualitative and quantitative research". *Behavioral and Brain Sciences*, vol. 45.

Burt, R. (1982). *Toward a structural theory of action. Network models of social structure, perception and action*. New York: Academic Press.

Curnow, R.; McLean, M. y Shepherd, P. (1976). *Techniques for analysis of system structures. SPRU Occasional papers series, n° 1*. Inglaterra: University of Sussex.

Danermark, B.; Ekström, M.; Jakobsen, L. y Karlsson, J. C. (2002). *Explaining Society. Critical Realism in the Social Sciences*. Londres: Routledge.

Dhir, S. y Singh, S. (2021). "Modified total interpretive structural modelling of innovation implementation antecedents". *International Journal of Productivity and Performance Management*, vol. 71, n° 4.

Faisal, M. N. (2010). "Analysing the barriers to corporate social responsibility in supply chains: an interpretive structural modelling approach". *International Journal of Logistics: Research and Applications*, vol. 13, n° 3, pp. 179-195.

Fatma, N.; Haleem, A.; Khan, M. I. y Khan, S. (2022). "Modelling of Barriers Towards the Adoption of Strategic Entrepreneurship: An Indian Context". En Faghih, N. y Forouharfar, A. (eds.), *Contextual Strategic Entrepreneurship . Contributions to Management Science*. Cham: Springer.

Gandhi, A. (2015). "Critical Success Factors in ERP Implementation and their Interrelationship Using MEIT and MICMAC". *Article in Indian Journal of Science and Technology*, vol. 8, n° S6, pp. 138-150.

Gardas, B.; Narkhede, B. y Raut, R. (2017). "A state-of the-art survey of interpretive structural modelling methodologies and applications". *International Journal of Business Excellence*, vol. 11, n° 4, pp. 505-560.

Glaser, B. G. y Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. New York: Aldine.

Gras, N., Dutrénit, G., & Vera-Cruz, M. (2019). A causal model of inclusive innovation for healthcare solutions: a methodological approach to implement a new theoretical vision of social interactions and policies. *Innovation and Development, 9*(2), 261-286.

Haleem, A.; Husain, Z. y Khan, J. (2017). "Barriers to technology transfer: a total interpretative structural model approach". *International Journal of Manufacturing Technology and Management*, vol. 31, nº 6, pp. 511-536.

Haleem, A.; Kumar, S. y Luthra, S. (2018). "Flexible System Approach for Understanding Requisites of Product Innovation Management". *Global Journal of Flexible Systems Management*, vol. 19, nº 1, pp. 19-37.

Jena, J.; Kumar Pathak, D.; Pandey, V.; Sidharth, S. y Thakur, L. (2017). "Total Interpretive Structural Modeling (MEIT): approach and application". *Journal of Advances in Management Research*, vol. 14, nº 2, pp. 162-181.

Kumbhat, A. y Sushil (2022). "Interactive Effect of Success Factors for High-Tech Startups: Value Propositions. Target Market and Operational Excellence". *International Journal of Global Business and Competitiveness (JGBC)*, vol. 17, pp. 73-88.

López-Castro, S. (2019). *La transferencia de conocimiento entre el Tecnológico Nacional de México y la industria: osbtáculos y su contexto organizacional*. Tesis de doctorado. España, Universidad Complutense de Madrid.

Martínez, N. (2014). *Relaciones estructurales y causalidad detrás de una innovación inclusiva: un caso de telemedicina en México*. Tesis de grado. México, DF, Universidad Autónoma Metropolitana Unidad Xochimilco.

Martinez, N; Dutrenit, G.; Gras, N.; y Tecuanhuey, E. (2018). "Actores, interacciones y causalidad en la innovacion inclusiva". *Innovar*, vol. 28, nº 70, pp. 23-38.

Maxwell, J. A. (1996). *Qualitative Research Design. An Interactive Approach. Applied Social Research Methods Series*, vol. 41. Thousand Oaks, California: SAGE.

_____ (2004). "Using Qualitative Methods for Causal Explanation". *Field Methods*, vol. 16, nº 3, pp. 243-264.

_____ (2021). "The importance of qualitative research for investigating causation". *Qualitative Psychology*, vol. 8, nº 3.

Menon, S. y Suresh, M. (2020). "Total Interpretive Structural Modelling: Evolution and Applications". En Bashar, A.; Raj, J. S. y Ramson, S. R. J. (eds.), *Innovative Data Communication Technologies and Application*, pp. 257-265. Cham: Springer.

Mukunda Das, V.; Sharma, A.; Singh, S. y Sinha, S. (2019). "A framework for linking entrepreneurial ecosystem with institutional factors: a modified total interpretive structural modelling approach". *Journal Global Business Advancement*, vol. 12, nº 3, pp. 382-404.

Pearl, J. (1995). "Causal diagrams for empirical research". *Biometrika*, vol. S2, nº 4, pp. 669-710.

_____ (2009). *CAUSALITY: Models, Reasoning, and Inference*, pp. 463. Los Ángeles: University of California.

Ruffa, C., & Evangelista, M. (2021). Searching for a middle ground? A spectrum of views of causality in qualitative research. *Italian Political Science Review/Rivista Italiana di Scienza Politica, 51*(2), 164-181.

Salmon, W. C. (1984). *Scientific Explanation and the Causal Structure of the World*. Princeton: Princeton University Press.

Strauss, A. L. (1987). Qualitative Analysis for Social Scientists. Cambridge: Cambridge University Press.

Sushil (2012). "Interpreting the Interpretive Structural Model". *Global Journal of Flexible Systems Management*, vol. 13, nº 2, pp. 87-106.

_____ (2017). "Modified ISM/TISM Process with Simultaneous Transitivity Checks for Reducing Direct Pair Comparisons". *Global Journal of Flexible Systems Management*, vol. 18, nº 4, pp. 331-351.

_____ (2018). "How to check correctness of total interpretive structural models?". *Annals of Operations Research*, vol. 270, nº 1-2, pp. 473-487.

_____ (2020). Interpretive multi-criteria ranking of production systems with ordinal weights and transitive dominance relationships. *Annals of Operations Research, 290*(1-2), 677-695.

von Wright, G. H. (1987). *Explicación y comprensión*. Madrid: Alianza.

Warfield, J. (1974). "Developing Interconnection Matrices in Structural Modeling". *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 4, n° 1, pp. 81-87.

Yin, R. (2003). *Case study research: Design and methods*. Thousand Oaks, CA: SAGE.

La colección **Ciencia, innovación y desarrollo** se propone reunir la producción académica relacionada con las ciencias básicas y aplicadas, el desarrollo tecnológico, la innovación, el emprendedurismo y el desarrollo.

La presente obra compila una serie de herramientas metodológicas, métodos y metodologías empleadas para el estudio de los procesos de CTI, con foco en las particularidades y adaptaciones necesarias para su uso en América Latina.

En este marco, los tres volúmenes constituyen un conjunto no exhaustivo de métodos y metodologías que han probado ser útiles para el estudio de estos procesos en la región, para un conjunto también no exhaustivo de tópicos sobre los cuales existe un intenso debate académico y de política pública.

El volumen 3 compila ocho capítulos en los que se incluyen técnicas mixtas y emergentes. Algunas de esas técnicas presentan fuertes solapamientos y articulaciones con los capítulos de los volúmenes anteriores, pero su tratamiento conceptual por separado contribuye a identificar fortalezas, debilidades y especificidades de cada una de las técnicas cuando se aplican a casos localizados en los países de América Latina.

Universidad Nacional de General Sarmiento

L U A
Libro Universitario Argentino

ISBN 978-987-630-748-2