



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**

**POSGRADO EN BIBLIOTECOLOGÍA Y ESTUDIOS DE LA INFORMACIÓN**

**FACULTAD DE FILOSOFÍA Y LETRAS  
INSTITUTO DE INVESTIGACIONES BIBLIOTECOLÓGICAS Y DE LA  
INFORMACIÓN**

***Metodología Neurocomputacional para el análisis de perfiles cuantitativos  
de la Ciencia Mexicana***

**TESIS**

**QUE PARA OPTAR POR EL  
GRADO DE: DOCTORA EN  
BIBLIOTECOLOGÍA Y ESTUDIOS  
DE LA INFORMACIÓN**

**PRESENTA:**

***Mtra. Ibis Anette Lozano Díaz***

**TUTOR PRINCIPAL:**

***Dr. Humberto A. Carrillo Calvet. Facultad de Ciencias***

**MIEMBROS DEL COMITÉ TUTOR:**

***Dr. Salvador Gorbea Portal. Instituto de Investigaciones Bibliotecológicas  
y de la Información***

***Dra. María Elena Luna Morales. CINVESTAV-IPN***

**Cd. Mx ., Agosto 2021**



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

***Dedicatoria:***

*A mis adorables padres, Ibís y Pedro; tenerlos aquí y allá es un privilegio. A Michael, con todo mi amor. A la memoria de mis abuelos.*

## Agradecimientos

Quiero agradecer a mi tutor **Dr. Humberto Carrillo Calvet** por su visión sobre el tema y por sus sabias sugerencias; su paciencia y exigencia en la **conducción de esta investigación**. Además, doy gracias por enseñarme el verdadero rigor de la ciencia, y cómo manejar cada situación que se presenta en el camino de hacer ciencia. A su lado, cada día es una enseñanza.

**A los miembros del comité tutor**, Dr. Salvador Gorbea Portal, Dra. María Luna Morales, Dr. Miguel Ángel Pérez Angón y Dr. Ricardo Arencibia Jorge por dedicar parte de su tiempo a la revisión de esta investigación y por sus valiosos comentarios, que sirvieron para enriquecer esta tesis. A la Dra. Yohannis Martí Lahera, directora de la Biblioteca Central de la Universidad de La Habana, por sus valiosos comentarios sobre la redacción de la tesis.

A mi colega y amigo José Luis Jiménez Andrade, porque sus comentarios a lo largo de la investigación doctoral han hecho una diferencia grande en la calidad del trabajo.

A mi formación ha contribuido de manera muy importante ser miembro del **Laboratorio de Dinámica no Lineal de la Facultad de Ciencias**. Lugar donde aprendí a trabajar con un equipo multidisciplinar; a recibir críticas; y al mismo tiempo hacer cuestionamientos.

**A los profesores del Posgrado en Bibliotecología y Estudios de la Información**, de la Universidad Nacional Autónoma de México por ofrecerme una excelente educación. Por último, agradecer al **CONACYT** su apoyo económico durante el período 2017-2021.

En el ámbito personal tengo que agradecer a mi familia, principalmente a mis padres, porque me han ayudado a avanzar en momentos que se sentían abrumantes. A Michael Moreno Jr., que me ha tenido una paciencia infinita. A Laura Moreno y Michael Moreno Sr. por quererme y apoyarme como una hija más.

## Resumen

El procesamiento y análisis de datos multidimensionales, para convertirlos en conocimiento útil (inteligencia), constituye un desafío para los analistas de las métricas de la información. Hoy en día, los recursos computacionales tienen limitaciones en el análisis y visualización de patrones en datos multidimensionales o para identificar perfiles cuantitativos multidimensionales similares.

Es por ello, que esta investigación se ha trazado como objetivo diseñar y aplicar una metodología orientada al proceso de análisis métrico de la información y la visualización de datos multidimensionales temporales obtenidos de bases digitales de artículos de investigación científica, aplicando técnicas de análisis bibliométrico y tecnologías basadas en inteligencia artificial. Esta metodología se ha denominado Mapeo Neurocomputacional de la Ciencia. Se ha incluido el esquema de etapas y tareas para el análisis y representación visual de los resultados.

Se ha utilizado el enfoque multidimensional para analizar los perfiles de desempeño determinados por los indicadores cuantitativos de productividad, impacto, excelencia, colaboración internacional, liderazgo e indicadores socioeconómicos. Para cada unidad de análisis, este conjunto de indicadores determina un perfil de desempeño multidimensional cuantitativo.

La metodología diseñada se aplica para analizar y comparar la dinámica de los perfiles de desempeño multidimensionales de la producción científica de México en el *Web of Science*, durante 2000 - 2019. Los análisis bidimensionales se realizaron con Bibliometrix de la librería de R. ViBlioSOM 2.0, como herramienta de *science mapping* permitió obtener varios tipos de mapas los cuales visualizan los perfiles cuantitativos multidimensionales y su evolución.

Los mapas multidimensionales permitieron comparar a México con países latinoamericanos; identificar el conjunto de países con similar desarrollo económico-demográfico y luego contrastar sus perfiles cuantitativos. Así como, comparar el perfil de las instituciones mexicanas más productivas. El mapeo neuro longitudinal facilitó el análisis de la evolución del perfil cuantitativo de las áreas de investigación para los años 2017, 2018 y 2019. El mapa orbital logró visualizar la dinámica del perfil cuantitativo de México en el período 2000-2019.

Esta metodología puede establecerse como modelo de referencia para realizar el análisis métrico de forma automatizada.

Palabras clave: cuantimetría, perfiles multidimensionales, inteligencia artificial, *science mapping*, México

## Tabla de Contenido

<b>Introducción</b> .....	<b>12</b>
<b>Exposición de motivos</b> .....	<b>12</b>
<b>Antecedentes y justificación</b> .....	<b>14</b>
<b>Problema de investigación</b> .....	<b>16</b>
<b>Objetivos de la investigación</b> .....	<b>16</b>
<b>Alcance del estudio</b> .....	<b>17</b>
<b>Estructura del documento</b> .....	<b>17</b>
<b>Capítulo 1: Marco teórico</b> .....	<b>18</b>
<b>1.1. Science of Science: un nuevo paradigma</b> .....	<b>18</b>
1.1.1. Science of Science y Ciencimetría: una relación virtuosa .....	19
1.1.2. La Ciencia como sistema complejo: retos de medición y análisis .....	21
<b>1.2. Descubrimiento de conocimiento en bases de datos (KDD): la perspectiva     cienciométrica</b> .....	<b>23</b>
1.2.1. Mapeo científico y visualización de información .....	25
1.2.2. Técnicas de mapeo científico.....	27
1.2.2.1. Análisis de citas .....	27
1.2.2.2. Acoplamiento bibliográfico .....	29
1.2.2.3. Análisis de co-citación.....	29
1.2.2.4. Análisis de co-ocurrencia de términos .....	31
1.2.3. Sistemas de software para la visualización .....	32
1.2.3.1. Redes de cocitación y colaboración.....	33
1.2.4. Inteligencia artificial en estudios cienciométricos.....	35
1.2.4.1. Mapas auto-organizados SOM .....	36
1.2.4.2. ViBlioSOM 2.0 – Software.....	37
1.2.5. Perfiles cienciométricos multidimensionales .....	39
<b>1.3. Indicadores, herramientas bibliométricas y cienciométricas</b> .....	<b>40</b>
1.3.1. Herramientas bibliométricas del Web of Science.....	41
1.3.1.1. InCites Benchmarking & Analytics: sus clases de métricas .....	42
1.3.1.2. Essential Science Indicators (ESI).....	43
1.3.1.3. Journal Citation Reports (JCR) .....	44
1.3.1.4. Esquemas de clasificación en la plataforma Web of Science .....	45
1.3.2. Indicadores cienciométricos.....	46
1.3.2.1. Indicadores dependientes e independientes de tamaño.....	46
1.3.2.2. Indicadores normalizados .....	47
<b>1.4. Metodologías y modelos para estudiar la ciencia</b> .....	<b>48</b>
<b>Capítulo 2: Metodología de la Investigación</b> .....	<b>50</b>
<b>2.1. Introducción</b> .....	<b>50</b>
<b>2.2. Metodología</b> .....	<b>50</b>
2.2.1. Etapa 1. Especificación de la investigación. ....	50
2.2.2. Etapa 2. Recuperación de datos y/o registros bibliográficos. ....	51
2.2.3. Etapa 3. Análisis biblio-infor-cienciométrico.....	51
2.2.3.1. Indicadores de producción y productividad .....	52

2.2.3.2. Indicadores de impacto y visibilidad real.....	53
2.2.3.3. Colaboración científica .....	54
2.2.3.4. Indicadores económicos-demográficos.....	54
2.2.3.5. Indicadores de innovación, liderazgo y actividad científica.....	55
2.2.4. Etapa 4. Análisis neurocomputacional y visualización.....	56
<b>Capítulo 3. Análisis y discusión de los resultados .....</b>	<b>57</b>
<b>3.1. Introducción.....</b>	<b>57</b>
<b>3.2. Mapeo Neurocomputacional de la Ciencia .....</b>	<b>57</b>
3.2.1. Etapa 1: Especificación de la investigación .....	58
3.2.2. Etapa 2. Recuperación de datos y/o registros bibliográficos .....	59
3.2.3. Etapa 3. Análisis biblio-infor-cienciométrico.....	59
3.2.4. Etapa 4: Análisis neurocomputacional y visualización.....	61
<b>3.3. Comprobación de la Metodología: Mapeo Neurocomputacional de la Ciencia. Su aplicación a la ciencia mexicana. ....</b>	<b>65</b>
<b>3.4. Mapeo de la evolución de la producción científica mexicana durante el período 2000-2019. ....</b>	<b>66</b>
3.4.1. Índice de productividad nacional.....	69
3.4.2. México en el contexto Latinoamericano.....	70
3.4.3. México en el contexto mundial.....	74
<b>3.5. Mapeo de los perfiles de desempeño cienciométrico de las instituciones mexicanas. ....</b>	<b>76</b>
3.5.1. Colaboración internacional.....	76
3.5.2. Producción e impacto de las instituciones más productivas .....	78
3.5.3. Impacto esperado y alto desempeño de las instituciones mexicanas .....	81
3.5.4. Comparación de los perfiles cienciométrico multidimensionales de las instituciones mexicanas .....	82
<b>3.6. Mapeo de la evolución de las áreas de investigación de la ciencia mexicana. ....</b>	<b>84</b>
3.6.1. Participación de producción en cada área de investigación de la producción nacional.....	84
3.6.2. Índice de actividad científica .....	84
3.6.3. Mapeo neuro longitudinal de las áreas de investigación: 2017- 2019 .....	86
<b>3.7. Mapeo de la evolución del perfil cienciométrico de México. ....</b>	<b>89</b>
<b>Capítulo IV. Consideraciones finales .....</b>	<b>91</b>
<b>Referencias .....</b>	<b>93</b>
<b>Anexos .....</b>	<b>114</b>
Anexo 1: Lista de acrónimos de las 60 instituciones mexicanas más productivas en WoS .....	114
Anexo 2: Lista de acrónimos de las instituciones extranjeras con mayor número de publicaciones en colaboración con México. ....	116
Anexo 3: Producción en WoS, Población total en 2019, Producción per cápita, Category Normalized Citation Impact, % %Documentos en el Top 1%, %Documentos en el Top 10%, % Colaboración internacional, % Documentos publicados en Q1y el % de Autores para Correspondencia de los 10 países Latinoamericanos más productivos en WoS. ...	118
Anexo 4: México y 29 países con similar desempeño socioeconómico: Población total en 2019, Producto Interno Bruto en 2019, Global Competitiveness Index 2019, Economic Complexity Index 2018.....	119
Anexo 5. Producción en WoS, Población total en 2019, Producción per cápita, Category Normalized Citation Impact, % %Documentos en el Top 1%, %Documentos en el Top	

10%, % Colaboración internacional, % Documentos publicados en Q1y el % de Autores para Correspondencia para México y 28 países.....	120
Anexo 6. Indicadores cientométricos que caracterizan el perfil de desempeño de las instituciones mexicanas más productivas en el período 2015-2019 en WoS. ....	122
Anexo 7: Perfil cientométrico de las 22 áreas de investigación mexicana en el WoS para el año 2017 .....	125
Anexo 8: Perfil de desempeño temático de las 22 áreas de investigación mexicana en el WoS para el año 2018 .....	126
Anexo 9: Perfil de desempeño temático de las 22 áreas de investigación mexicana en el WoS para el año 2019 .....	127
Anexo 10: Valores anuales de los indicadores que caracterizan el perfil cientométrico de México en WoS entre 2000-2019.....	128

## Relación de Figuras

Figura 1. Fases de la metodología para el descubrimiento de conocimiento en bases de datos (KDD)(Fayyad, Piatetsky-Shapiro, & Smyth, 1996).....	24
Figura 2. Estructura de una Red Neuronal bajo el algoritmo SOM .....	36
Figura 3. InCites Benchmarking & Analytics (InCites B&A).....	41
Figura 4. Essential Science Indicators (ESI) .....	43
Figura 5. Journal Citation Reports (JCR) .....	44
Figura 9. Notebook de la producción científica mexicana en GoogleColab, programado en pyhton y utilizando plotly. ....	56
Figura 6. Modelo de la metodología. ....	58
Figura 7. Análisis biblio-infor-cienciométrico de la producción científica con Bibliometrix R-package. ....	61
Figura 8. Pasos para el procesamiento neurocomputacional de datos multidimensionales que caracterizan el perfil cienciométrico de un país, institución o campo de conocimiento. ....	62
Figura 10. Evolución de la producción de México en Web of Science y Scopus, 2000-2019.....	66
Figura 11. Comparación de la dinámica de las tasas de crecimiento de la producción científica, el producto interno bruto y la población en México durante 2000-2019.....	67
Figura 12. Dinámica de los tipos de producción científica mexicana en Web of Science, durante 2000-2019. ....	68
Figura 13. Dinámica del total de artículos mexicanos publicados en revistas nacionales y extranjeras indexadas en el WoS, durante el 2000-2019.....	68
Figura 14. Evolución de miembros del Sistema Nacional de Investigadores y la producción total en WoS. ....	69
Figura 15. Evolución de la productividad nacional mexicana por tipo de producción en WoS, 2000-2019.....	70
Figura 16. Evolución del volumen de artículos en WoS por cada 10,000 habitantes para diez países Latinoamericanos más productivos.....	71
Figura 17. Comparación entre la producción por cada país de Latinoamerica como porcentaje de la producción mundial y el impacto normalizado (2000-2019) .....	71
Figura 18. Relación entre la producción científica de excelencia y el liderazgo latinoamericano en la investigación, durante 2000-2019. El tamaño de las bolas representa el número de artículos en el período.....	72
Figura 19. Perfil cienciométrico multidimensional de los 10 países latinoamericanos más productivos en WoS , durante el período 2017-2019. ....	73
Figura 20. Comparación de 108 países utilizando indicadores económico-demográficos. Los códigos de país utilizan el estándar internacional ISO 3166-1 alpha-3.....	75
Figura 21. Perfiles cienciométricos multidimensionales de países con perfiles económicos-demográficos más similares a México (2017-2019). ....	76

Figura 22. Instituciones y países con mayor colaboración científica con México en el WoS, 2000-2019. Los acrónimos de las instituciones se pueden consultar en el Anexo 2. ....	77
Figura 23. Impacto de la investigación mexicana con instituciones extranjeras, 2000-2019.....	78
Figura 24. Producción e impacto de las 60 instituciones mexicanas más productivas en WoS, durante el 2015-2019.....	79
Figura 25. Producción y productividad(Institutional Scientific Productivity) de las 60 instituciones mexicanas más productivas en WoS, durante el 2015-2019. ....	80
Figura 26. Impacto normalizado (Category Normalized Citation Impact) versus la productividad institucional (ISP) durante 2015-2019.....	81
Figura 27. Producción de alto desempeño y visibilidad esperada en 60 instituciones mexicanas más productivas en WoS, durante 2015-2019.....	82
Figura 28. Perfil multiparamétrico de las 60 instituciones mexicanas más productivas en WoS, durante el período 2015-2019. ....	83
Figura 29. Participación de la producción académica mexicana en revistas con factor de impacto, 2000–2019. ....	84
Figura 30. Revealed Comparative Advantage para las áreas de investigación mexicanas en 2000, 2005, 2010, 2015 y 2019.....	85
Figura 31. Revealed Comparative Advantage para México en 22 campos científicos en 2005 y 2019. El país es comparado con el resto del mundo. El tamaño de la bola está representado por la contribución de México en cada campo en la producción nacional en el 2019.....	86
Figura 32. Mapeo neurolongitudinal de las áreas de investigación mexicanas en el Essential Science Indicators, 2017-2019. ....	88
Figura 33. Mapeo de la evolución del perfil de desempeño cientiométrico de México en WoS, 2000-2019.....	90

## Glosario de abreviaturas de países

	<b>Países</b>	<b>Siglas</b>		<b>Países</b>	<b>Siglas</b>
1	Albania	ALB	37	Ecuador	ECU
2	Algeria	DZA	38	Egypt, Arab Rep.	EGY
3	Angola	AGO	39	El Salvador	SLV
4	Argentina	ARG	40	Estonia	EST
5	Armenia	ARM	41	Ethiopia	ETH
6	Australia	AUS	42	Finland	FIN
7	Austria	AUT	43	France	FRA
8	Azerbaijan	AZE	44	Gabon	GAB
9	Bahrain	BHR	45	Gambia, The	GMB
10	Bangladesh	BGD	46	Georgia	GEO
11	Barbados	BRB	47	Germany	DEU
12	Belgium	BEL	48	Ghana	GHA
13	Benin	BEN	49	Greece	GRC
14	Bolivia	BOL	50	Guatemala	GTM
15	Bosnia and Herzegovina	BIH	51	Guinea	GIN
16	Botswana	BWA	52	Haiti	HTI
17	Brazil	BRA	53	Honduras	HND
18	Brunei Darussalam	BRN	54	Hong Kong SAR, China	HKG
19	Bulgaria	BGR	55	Hungary	HUN
20	Burkina Faso	BFA	56	Iceland	ISL
21	Burundi	BDI	57	India	IND
22	Cabo Verde	CPV	58	Indonesia	IDN
23	Cambodia	KHM	59	Iran, Islamic Rep.	IRN
24	Cameroon	CMR	60	Ireland	IRL
25	Canada	CAN	61	Israel	ISR
26	Chad	TCD	62	Italy	ITA
27	Chile	CHL	63	Jamaica	JAM
28	China	CHN	64	Japan	JPN
29	Colombia	COL	65	Jordan	JOR
30	Congo, Dem. Rep.	COD	66	Kazakhstan	KAZ
31	Costa Rica	CRI	67	Kenya	KEN
32	Cote d'Ivoire	CIV	68	Kuwait	KWT
33	Cyprus	CYP	69	Kyrgyz Republic	KGZ
34	Czech Republic	CZE	70	Lao PDR	LAO
35	Denmark	DNK	71	Latvia	LVA
36	Dominican Republic	DOM	72	Lebanon	LBN

	<b>Países</b>	<b>Siglas</b>		<b>Países</b>	<b>Siglas</b>
73	Lesotho	LSO	111	Slovak Republic	SVK
74	Lithuania	LTU	112	Slovenia	SVN
75	Luxembourg	LUX	113	South Africa	ZAF
76	Malawi	MWI	114	Spain	ESP
77	Malaysia	MYS	115	Sri Lanka	LKA
78	Mali	MLI	116	Sweden	SWE
79	Malta	MLT	117	Switzerland	CHE
80	Mauritania	MRT	118	Tajikistan	TJK
81	Mauritius	MUS	119	Tanzania	TZA
82	Mexico	MEX	120	Thailand	THA
83	Moldova	MDA	121	Trinidad and Tobago	TTO
84	Mongolia	MNG	122	Tunisia	TUN
85	Montenegro	MNE	123	Turkey	TUR
86	Morocco	MAR	124	Uganda	UGA
87	Mozambique	MOZ	125	Ukraine	UKR
88	Namibia	NAM	126	United Arab Emirates	ARE
89	Netherlands	NLD	127	United Kingdom	GBR
90	New Zealand	NZL	128	United States	USA
91	Nicaragua	NIC	129	Uruguay	URY
92	Nigeria	NGA	130	Venezuela, RB	VEN
93	Norway	NOR	131	Vietnam	VNM
94	Oman	OMN	132	Yemen, Rep.	YEM
95	Pakistan	PAK	133	Zambia	ZMB
96	Panama	PAN	134	Zimbabwe	ZWE
97	Paraguay	PRY			
98	Peru	PER			
99	Philippines	PHL			
100	Poland	POL			
101	Portugal	PRT			
102	Qatar	QAT			
103	Romania	ROU			
104	Russian Federation	RUS			
105	Rwanda	RWA			
106	Saudi Arabia	SAU			
107	Senegal	SEN			
108	Serbia	SRB			
109	Seychelles	SYC			
110	Singapore	SGP			

## Introducción

### Exposición de motivos

La ciencia ha sido analizada, a través de los siglos, por numerosos filósofos, matemáticos, físicos, historiadores y sociólogos, y desde múltiples perspectivas (Carnap, 1995; Kuhn, 1996; Selin, 2006; Yuexiao, 1988). Ha experimentado momentos de negación y represión, de independencia e incuestionabilidad, y de crítica desde la diversidad. Desde las posiciones mertonianas de la Ciencia como método certificador del conocimiento, acervo de conocimiento acumulado, conjunto de valores y normas culturales que gobiernan las actividades científicas, o como combinación de todos los elementos anteriores (Merton, 1942; Turnpenny, Jones, & Lorenzoni, 2011; Ziman, 2000) hasta la visión constructivista y antagónica de Bruno Latour (1979), evaluadora de los objetos de estudio científico como constructos sociales permeados de decisiones subjetivas en torno a la validez o no de los datos obtenidos; la ciencia ha sido definida y contextualizada como sistematización lógico-matemática, mediante el uso de la teoría general del símbolo, el lenguaje artificial y la semiótica, lo cual ha sido considerado por (Geisler, 2005) como una visión muy limitada.

En la presente investigación la ciencia se define como una red social de investigadores que generan y validan una red de conocimiento. Se centra en dos visiones fundamentales: actividad social y red de conocimiento; sobre la base de elementos de la teoría normativa (Merton, 1942). La visión de la ciencia como actividad social es un análisis de las interacciones entre académicos y las instituciones a la cual pertenecen para generar nuevo conocimiento. La ciencia como red de conocimiento es concebida como una estructura cognitiva que consiste en las relaciones entre las ideas, las cuales, están plasmadas en revistas y documentos científicos (Scharnhorst, Borner, & van den Besselaar, 2012).

La “ciencia moderna” se ha convertido en objeto de estudio debido a la generación constante de grandes volúmenes de datos, que son almacenados en un conjunto de bases de datos de extraordinaria importancia para la actividad científica, como son las bases de datos Scopus, Google Scholar, *Web of Science*, PubMed, *Microsoft Academic*, entre otras. Cada una de estos índices brinda facilidades para que un investigador, en el menor tiempo posible, sea capaz de encontrar, entre todo el cúmulo de datos o información disponible, aquella que sea pertinente a su objeto de investigación. Además, cada vez con más frecuencia estos datos son analizados desde la visión multidisciplinar de científicos que se desenvuelven en diversos dominios del conocimiento. En este escenario, emerge un campo interdisciplinario para el estudio y predicción del conocimiento científico: “*Science of Science*”, disciplina definida por primera vez por el filósofo John Bernal en el XI Congreso Internacional de Historia de la Ciencia, celebrado en Varsovia en 1965 (Bernal & Mackay, 1966). Paralelamente, en la revista *Nature* en 1965, Maurice Goldsmith la definió como “el examen del fenómeno de la 'ciencia' empleando los métodos de la ciencia misma” (Goldsmith, 1965).

Katy Börner se refiere a “*Science of Science*” como el estudio científico de todas las actividades de investigación en ciencias naturales, sociales y artes/humanidades (Borner & Scharnhorst, 2009). *Science of Science* permite analizar la ciencia considerando datos estáticos (publicaciones científicas, redes de colaboración,

redes de citas), así como la dinámica de las publicaciones, el impacto científico, la predicción en la ciencia, la interdisciplinariedad, la innovación y la difusión del conocimiento (Garfield, 2009). Por tanto, un rasgo distintivo de esta disciplina es la integración transdisciplinaria de métodos y teorías de otras áreas científicas para estudiar la ciencia contemporánea; entendida como: interacciones complejas entre las estructuras sociales, las representaciones del conocimiento y el mundo natural (Fortunato et al., 2018).

En este sentido, estudiar la ciencia considerando elementos de las ciencias de la computación, matemáticas, sociología y ciencia de la información permite describir la evolución, regularidades y relaciones en las actividades de investigación, así como, el conjunto de productos que generan estas actividades.

*Science of Science* necesita de las plataformas de información y sus herramientas debido al enorme volumen de datos sobre la ciencia que hoy son almacenados diariamente. De la ciencia de la información adopta herramientas, métodos e indicadores procedentes de las especialidades métricas de la información: bibliometría, informetría y otra muy relacionada con estas como la cienciometría, para caracterizar las regularidades de la ciencia (Mingers & Leydesdorff, 2015; Pritchard, 1969; Sideris, Katsaros, Sidiropoulos, & Manolopoulos, 2018). Resulta conveniente precisar aquí, las tres dimensiones de la bibliometría: la descriptiva o clásica, la histórica y la evaluativa, esta última definida por Francis Narin como “el empleo de las técnicas bibliométricas, especialmente el análisis de las publicaciones y las citas, en la evaluación de la actividad científica”(Narin, 1976).

De la sociología, *Science of Science*, adopta teorías y procesos sociales; y de las ciencias de la computación incorpora algoritmos de inteligencia artificial, machine learning, y otras técnicas de mapeo auto-organizado (Self Organizing Maps), basadas en redes neuronales artificiales (Kangas, Kohonen, & Laaksonen, 1990), y que son útiles para descubrir nuevo conocimiento a partir de la minería de datos y el *Knowledge Discovery in Databases* (KDD) (Fayyad & Stolorz, 1997). Por último, de las matemáticas emplea modelos estadísticos y Agent-Based Models of Science (Payette, 2012) para modelar, entender y caracterizar la Ciencia como un sistema dinámico y complejo. También se basa en métodos de visualización de información (Boyack, Wylie, & Davidson, 2015) y science mapping (Chen, 2017) con el propósito de representar el nuevo conocimiento.

Desde este enfoque transdisciplinar, resulta factible la propuesta de una metodología que permita analizar, comparar y visualizar los perfiles multidimensionales de un dominio del conocimiento (Villaseñor, Carrillo-Calvet, & Arencibia-Jorge, 2017), sea cual fuere su nivel de agregación (temático, individual, institucional, nacional, regional, mundial), y caracterizar su evolución a partir de estudios longitudinales (Harzing & Alakangas, 2016). Esta propuesta se basa en la hipótesis de que, para comprender profundamente la ciencia, se debe abordar considerando elementos, fuentes, herramientas, teorías, y conceptos procedentes de varias disciplinas científicas, con el objetivo de capturar la estructura y la dinámica del esfuerzo científico para obtener información sobre el funcionamiento interno de la ciencia.

En este escenario, la Bibliotecología y la Ciencia de la Información demuestra su carácter interdisciplinar con otras áreas de conocimiento (Hjørland & Albrechtsen,

1995). Específicamente, las métricas de la información reciben la influencia de novedosos aspectos teóricos, conceptuales y prácticos de otras disciplinas (matemáticas, computación, tecnologías de la información, etc.) para describir y analizar los problemas relacionados con la caracterización y predicción del comportamiento de la actividad científica.

Los científicos/profesionales de la información, para llevar a cabo investigaciones exhaustivas sobre la estructura, la evolución y el contenido de la ciencia, emplean sofisticadas herramientas (software) (Börner, 2011; Chen, 2006a; Evans & Foster, 2011; Soós, 2011; Thor, Marx, Leydesdorff, & Bornmann, 2016; N. J. Van Eck & Waltman, 2014) que son excelentes para analizar grandes conjuntos de datos bibliométricos, y que requieren de una cada vez mayor interrelación con las Ciencias de la Computación, las Matemáticas, u otros dominios interdisciplinarios para representar el nuevo conocimiento. En este contexto, durante los últimos años se observa una evolución en las formas de realizar estudios métricos, que van más allá de la aplicación de un conjunto de indicadores tradicionales, y se enfocan en lograr una visión holística e integradora a la hora de analizar un dominio del conocimiento (Hjørland, 2002)

Se espera que estudiar la ciencia desde una perspectiva transdisciplinar, con el uso de sus propios métodos científicos, permita formular políticas científicas: en el gobierno, en los consejos de investigación, agencias encargadas de establecer programas y fondos de investigación en universidades y centros de investigación, y otros sectores. Además, este enfoque debe facilitar la supervisión y previsión en la dinámica de los avances científicos y tecnológicos, y permitir una mejor comprensión de las peculiaridades institucionales para alcanzar un sistema de investigación de alto rendimiento. Por último, es la oportunidad para investigadores, organizaciones de investigación y financiadores, de mostrar el impacto social que están teniendo sus investigaciones.

### **Antecedentes y justificación**

Entre los esfuerzos por normalizar y estandarizar el análisis métrico, en el contexto de la evaluación de la investigación, aparecen experiencias que tratan en profundidad la aplicación de indicadores y el modo de obtenerlos. Una de estas experiencias, está recogida en los manuales de la Organización para la Cooperación y el Desarrollo Económico (OECD/Eurostat, 2015), que definen indicadores de innovación y producción científica; aunque el principal objetivo en ellos es facilitar la evaluación e interpretación de los datos sobre investigación y desarrollo (I+D) obtenidos.

En la región latinoamericana, se creó en abril de 1995, la Red Iberoamericana e Interamericana de Indicadores de Ciencia y Tecnología (RICYT). Con ella se implementó una plataforma para monitorear, a través de indicadores científicos, el desempeño de la ciencia y la tecnología en la región Iberoamericana (RICYT, 2019).

Particularmente, la ciencia mexicana en las últimas décadas ha sido estudiada desde diferentes perspectivas. Por ejemplo, el proyecto cuantitativo sobre el Atlas de la Ciencia Mexicana (Perez-Angón, 2018) destaca por analizar la ciencia mexicana combinando las bases de datos del Sistema Nacional de Investigadores, el *Web of Science* y Scopus. El reporte “Principales indicadores cuantitativos de la producción mexicana, 2018” presenta las principales tendencias de la actividad

científica mexicana, considerando indicadores de producción, visibilidad, impacto, colaboración, liderazgo y excelencia entre 2008-2017 en la base de datos Scopus (Elsevier), y el *SCImago Research Group* aplicó sus métodos en el procesamiento, normalización y visualización de los datos (SCIMAGO, 2018). Otro estudio cuantitativo muy reciente de la producción científica mexicana se realizó en Scopus (Lancho-Barrantes & Cantú-Ortiz, 2019), determinando los países temáticamente homólogos con México. Iniciativas que, su objetivo principal es describir la ciencia mexicana aplicando técnicas e indicadores bibliométricos.

En los últimos veinte años se han publicado estudios bibliométricos dedicados al análisis de diversos dominios del conocimiento en la Ciencia mexicana, en los cuales se ha utilizado la información contenida en Scopus, *Web of Science* y otras bases de datos nacionales como REDALYC, Clase y Periódica. La investigación más reciente analiza la evolución de los artículos publicados por instituciones mexicanas en el campo de la Física Aplicada en *Web of Science*, en un período de 45 años (del Río, Russell, & Juárez, 2020). Dominios como la física, astronomía, salud, la psicología (Arvanitis, Russell, & Rosas, 1996; Collazo-Reyes, Luna-Morales, Russell, & Perez-Angon, 2008; Lena, 1997; Luna Morales, Luna Morales, & Sanchez Martinez, 2013), entre otros, también han sido caracterizados con el uso de indicadores bibliométricos (Bravo Vinaja, 2005; Castillo-Pérez & Muñoz-Valera, 2014; de Arenas, Castanos-Lomnitz, & Arenas-Licea, 2002; Gantman & Fernandez Rodriguez, 2017; Luis Alexandre, Luis Alexandre-Tudo, Bolanos-Pizarro, & Alexandre-Benavent, 2016; Perez-Santos & Anaya-Ruiz, 2013; Sierra-Flores, Guzmán, Raga, & Pérez, 2009).

También aparecen estudios sobre la región latinoamericana en los cuales se exponen algunas características del desempeño cuantitativo de México en comparación con Brasil, Chile, Argentina entre otros (Alcazar Farias & Lozano Guzman, 2009; Chinchilla-Rodriguez, Zacca-Gonzalez, Vargas-Quesada, & Moya-Aregon, 2015; Collazo-Reyes, Luna-Morales, & Luna-Morales, 2017; Macias-Chapula, 2013).

Un antecedente muy cercano a la presente investigación es la propuesta metodológica para identificar y describir la investigación mexicana sobre química en bases de datos bibliográficas (Russell, Hernandez-Garcia, & Kleiche-Dray, 2016). Esta metodología permitió recuperar datos sobre química en los sistemas, y conformar una base de datos bibliométrica para analizar este campo.

Entre las iniciativas antes mencionadas, y en una amplia revisión de la literatura sobre el tema, no se encontró una metodología enfocada a la visualización, caracterización y comparación de perfiles multidimensionales y su evolución, aprovechando las capacidades de la red neuronal artificial (RNA) conocida como SOM (por las siglas en inglés de *Self-Organizing Maps*).

La metodología que se presenta en esta investigación tiene como antecedente el desarrollo de ViBlioSOM, metodología orientada al proceso de análisis métrico de la información y la visualización de matrices de co-ocurrencia en forma de mapas auto-organizados (Guzmán Sánchez, 2009). Esta propuesta proviene del trabajo en equipo del Laboratorio de Dinámica no Lineal de la Facultad Ciencias de la Universidad Nacional Autónoma de México (UNAM), el cual ha venido desarrollando software basado en inteligencia computacional para ofrecer soluciones a problemas

vinculados a los estudios métricos de la información (Sotolongo Aguilar, Guzmán Sánchez, & Carrillo, 2002; Villaseñor et al., 2017).

Los desarrollos anteriores, contienen una descripción del proceso general, estableciendo un conjunto de tareas a seguir para realizar el análisis de información métrica, conjugado con el proceso de KDD. Este proceso necesita, no obstante, una nueva etapa de desarrollo: traspolar modelos y recursos matemáticos y computacionales para analizar, visualizar y comparar el estado actual y evolución de perfiles multidimensionales en la producción científica mexicana. Es por ello, que la autora se ha planteado:

### **Problema de investigación**

¿Cómo integrar diversas técnicas procedentes de la bibliometría, de las matemáticas, de la visualización de información, y de la inteligencia artificial, para diseñar una metodología que permita describir, analizar y representar visualmente la evolución de los perfiles cuantitativos multidimensionales de países, campos de investigación, instituciones académicas o investigadores?

### **Objetivos de la investigación**

#### **Objetivo General**

Diseñar una metodología que integre técnicas bibliométricas, matemáticas, de visualización de información y de inteligencia artificial para hacer minería de datos en las bases de datos de artículos científicos con el propósito de visualizar el conocimiento obtenido acerca de los perfiles cuantitativos multidimensionales de países, campo de investigación, institución académica o investigadores.

#### **Objetivos Específicos**

Determinar el marco-teórico referencial de la investigación, considerando las bases conceptuales y metodológicas que permitan analizar la compleja actividad científica y describir su evolución desde una perspectiva cuantitativa y multidimensional.

Determinar métodos y técnicas basadas en redes neuronales autorganizadas que permitan obtener y visualizar conocimiento en las bases de artículos de investigación científica para caracterizar la evolución del perfil cuantitativo de una institución, país o campo de conocimiento.

Aplicar la metodología obtenida en el contexto mexicano, para la caracterización del perfil cuantitativo nacional, sus principales áreas de investigación y sus instituciones líderes, minando las bases de datos del *Web of Science*.

## **Alcance del estudio**

El aporte y alcance de la presente investigación se lista a continuación:

1. Las bases teóricas de los mapeos de la ciencia (*science mapping*) con redes neuronales artificiales bajo el enfoque metodológico de la Bibliotecología y Ciencia de la Información.
2. Aplicación de redes neuronales artificiales para el análisis y visualización de datos multidimensionales cuantitativos.
3. Una metodología útil para el análisis y el mapeo auto-organizado de la evolución de perfiles multidimensionales en un dominio del conocimiento.

Los datos bibliográficos están limitados por la fuente de la recuperación, es decir, el *Web of Science*. La base de datos presenta deficiencias de cobertura reportadas en la literatura (baja cobertura de revistas latinoamericanas, exclusión de citas provenientes de revistas en idiomas distintos del inglés, entre otras), lo cual da un margen de error a los análisis de citas y a los estudios derivados (Chinchilla-Rodríguez et al., 2015).

Las irregularidades presentadas en los campos “afiliación institucional” y “autor” exigieron un riguroso trabajo de desambiguación y control de autoridades, en el volumen de información de *Web of Science* recuperado y procesado durante el período 2000-2019.

## **Estructura del documento**

La presente tesis queda estructurada a partir de: introducción, tres capítulos, consideraciones finales, bibliografía y anexos. La bibliografía que ha sido referenciada a partir del estilo bibliográfico de la *American Psychological Association* (APA 6<sup>ta</sup> Edición), y contiene varios anexos.

En el Capítulo I, se abordan los elementos teóricos-conceptuales que sustentan la propuesta, así como los más novedosos avances sobre el tema. El Capítulo II expone los aspectos metodológicos para analizar la producción científica de México en el *Web of Science*, durante 2000-2019. En el Capítulo III se presentan los resultados de la investigación: la metodología diseñada para describir, analizar y representar visualmente la evolución de los perfiles multidimensionales cuantitativos en la ciencia; y se analizan y discuten los resultados obtenidos al aplicar esta metodología en el estudio de caso enfocado en la producción científica de México en el *Web of Science*. Finalmente, en el Capítulo IV se presentan las consideraciones finales.

## Capítulo 1: Marco teórico

### 1.1. Science of Science: un nuevo paradigma

La ciencia ha sido objeto de estudio por más de cien años. En 1923 los editores de la revista *Nauka Polska* (Ciencia polaca), realizaron un estudio sobre los contenidos de los números anteriores publicados en ella. Esta investigación permitió crear una sección llamada “conocimientos sobre la ciencia”, dedicada a la creatividad científica y las condiciones para su desarrollo. A partir de este momento, identificaron por primera vez la necesidad de proponer un campo científico relacionado con los problemas de la ciencia (Skalska-Zlat, 2001).

Ciencia de la Ciencia o “*Science of Science*” (SoS) comenzó a definirse como disciplina autónoma a partir del año 1925 en Polonia. En este año el sociólogo polaco Florian Znaniecki (1881-1958) publicó su estudio *Przedmiot i zadania nauki o wiedzy* (Asunto y tareas de la ciencia del conocimiento), el cual sentó las bases metodológicas de la nueva disciplina. Entre 1928 y 1939 se organizaron más de cuarenta reuniones en las que reconocidos intelectuales polacos discutieron sobre la nueva disciplina y el término fue transformándose desde *wiedza* o *nauce* hasta *naukosnawstwo*, traduciendo como ciencia de la ciencia.

Diez años después en 1935, Maria (1896-1974) y Stanisław (1897-1963) Ossowski publicaron un artículo en la revista *Nauka Polska*, definiendo el alcance y formularon el programa de investigación ciencia de la ciencia. En este trabajo, el paradigma de ciencia de la ciencia involucraba los productos intelectuales de la ciencia y las actividades científicas como parte de la cultura (Sheehan, 2007). Así, científicos polacos prepararon el terreno para las futuras investigaciones en este campo.

La ocupación nazi en Polonia (1935) impidió el desarrollo de este paradigma hasta después de la Segunda Guerra Mundial. Sin embargo, intelectuales como el físico ruso Boris Hessen siguieron trabajando en este nuevo paradigma. La contribución de Hessen devino en el desarrollo de una ciencia de la ciencia esencialmente igual a la creada por los filósofos polacos (Wouters, 1999).

En 1939 John Desmond Bernal, publica *The Social Function of Science*, la cual resultó su obra maestra. Ignorando a los intelectuales polacos y rusos, en el libro Bernal no utiliza el término “ciencia de la ciencia”, pero aun así, es considerado como uno de los fundadores de este campo. Su libro contribuyó a la idea de analizar científicamente la ciencia y cuantificar el estado de la ciencia. Destacó la necesidad de un fundamento científico para la política científica, la aplicación sistemática de los métodos científicos (Sheehan, 2007) a la ciencia misma y la idea de una unidad general de la ciencia.

La obra de influyentes científicos estuvo influenciada por las ideas de Bernal: Eugene Garfield, Joshua Lederberg, Robert Merton, Derek Price y V. V. Nalimov. Particularmente, Derek Price fue fundamental para el movimiento de la ciencia de la ciencia. En 1950 en el *Congrès International d'Histoire des Sciences*, Price presentó su primer trabajo sobre medidas cuantitativas del desarrollo de la ciencia (Price, 1951). Este trabajo no tuvo respuesta de la comunidad científica porque su audiencia no estaba familiarizada con el uso de métodos cuantitativos basándose en la econometría y la termodinámica (Price, 1951). Este evento motivó a Price a escribir *Little Science, Big Science*, la cual publicó 1962 y fue su obra maestra.

Para Price el estudio de la ciencia debía permitir el diagnóstico de la naturaleza de los cambios que se están produciendo actualmente y planificar en consecuencia la disposición de las instalaciones de investigación en la universidad y laboratorios (Price, 1963). Con la creación del *Science Citation Index* (SCI), Price identificó la oportunidad de acceder a datos que le permitieron obtener conocimiento científico objetivo y sólido sobre la ciencia empleando métodos científicos (Wouters, 1999).

En 1965, Maurice Goldsmith primer director de la nueva Fundación Ciencia de la Ciencia publicó un artículo en *Nature* donde explicaba sus objetivos. En este artículo se define la ciencia de la ciencia como el examen del fenómeno de la 'ciencia' por los métodos de la ciencia misma (Goldsmith, 1965). Además, este tipo de estudio involucra la sociología de la ciencia, la psicología, investigación operativa sobre ciencia, la economía de la ciencia, la planificación de la ciencia y el análisis del flujo de información científica.

Influentes científicos en diferentes campos de investigación publicaron recientemente un artículo en la revista *Science*, donde retomaron con fuerza renovadora los viejos planteamientos de sus antecesores (Fortunato et al., 2018). En este importante trabajo se consideran las ideas de los polacos y rusos, al plantear que el principal valor de la ciencia de la ciencia es la comprensión profunda de los factores que pueden impulsar la ciencia hacia el éxito. De manera que los problemas ambientales, sociales y tecnológicos se pudieran abordar en la actualidad, con el despliegue multidisciplinar de todo el arsenal de herramientas analíticas disponible, y aprovechando la mayor velocidad de cómputo de los sistemas de procesamiento y análisis de información, con mucha mayor efectividad.

El paradigma SoS ha evolucionado en la medida que la ciencia se desarrolla. La ciencia puede describirse hoy como una red compleja, autoorganizada y en evolución de académicos, proyectos, artículos e ideas (Scharnhorst et al., 2012). En este escenario, SoS ofrece: una comprensión detallada de la estructura relacional entre científicos, instituciones e ideas; conocimientos sobre patrones pasados, lo que ayuda a guiar la investigación futura dentro de sus disciplinas y la modelación de oportunidades adaptadas a las características de cada dominio. Por tal motivo, SoS integra hallazgos y teorías de múltiples disciplinas: la cienciometría; estadísticas descriptivas; visualización de datos; enfoques de ciencia de redes; algoritmos de aprendizaje automático; análisis matemático y simulación por computadora (Fortunato et al., 2018).

### 1.1.1. Science of Science y Cienciometría: una relación virtuosa

Bernal en 1939 y Price en 1961, ejercieron una gran influencia en el auge de la ciencia de la ciencia, y por consiguiente, aparecen nuevas disciplinas. Esta cultura científica despertó el interés de otros científicos en el este europeo en la década de los años 60. Dobrov en Ucrania y Vassily Vassilievich Nalimov en Moscú fueron los principales proponentes de los estudios científicos cuantitativos cuando al mismo tiempo se estaba consolidando la ciencia de la ciencia. En esta sección, ilustraremos la relación entre la ciencia de la ciencia y la cienciometría.

La diversidad de pensamiento en ambos científicos dio lugar a diferentes enfoques de la cienciometría. Dobrov, fue en su generación el primero en publicar un libro sobre la ciencia de la ciencia en la entonces Unión Soviética. Promovió la

idea que, el campo de los estudios cuantitativos era fundamental para conducir la política científica. También propuso un término más coherente para denominar esta disciplina: Naukovedenie (en ruso) que traducido al inglés es *Scientology* y al español como Cienciología, pero esta denominación no ha sido muy utilizada debido a que tiende a confundirse con una secta religiosa que lleva esa misma denominación, en su lugar los rusos aceptaron más Nauka o nauke (Ciencia de la Ciencia).

En cambio, Nalimov entendía la ciencia como un sistema autoorganizado y destacó la necesidad de un sistema científico abierto. Para Nalimov los estudios científicos deberían mejorar la autocomprensión de la ciencia, pero nunca servir como objetivo instrumental (Wouters, 1999). Con el tiempo la obra de Nalimov tuvo un impacto mayor en la comunidad científica.

En 1966 más de cien académicos participaron en el Simposio Soviético-Polaco de "ciencia de la ciencia". Durante este evento la SoS se abordó como una disciplina compleja que se ocupaba de la interacción de varios elementos de la ciencia y el conocimiento sobre un área en su totalidad (Granovsky, 2001).

Nalimov planteó las principales tareas de la ciencia de la ciencia en este simposio: el estudio de la producción científica, el descubrimiento de optimizar la organización científica y lograr la efectividad en las actividades de investigación. Además, ofreció la solución a estas tareas mediante la combinación de los enfoques descriptivos y cuantitativos (Skalska-Zlat, 2001). Demostrando, la amplia aplicación de los métodos matemáticos en el campo de la ciencia de la ciencia.

La cristalización de estas ideas fue en 1969, cuando Vassily V. Nalimov y Z. M. Mulchenko publicaron el libro *Naukometrija. Izucente rezvitja nauki kak informacionnogo processa (Scientometrics. The study of science as an information process)* (Nalimov & Mulcsenko, 1969). En este libro acuñaron el equivalente ruso del término "cienciometría" ("naukometrija") para estudiar el desarrollo de la ciencia. Definiéndose como la "Aplicación de métodos cuantitativos a la investigación sobre el desarrollo de la ciencia como un proceso informativo"(Nalimov & Mulcsenko, 1969).

El trabajo abordó por primera vez los métodos cuantitativos en el estudio de la ciencia y amplió los principios definidos en trabajos anteriores. Debido al amplio reconocimiento de la cienciometría, Tibor Braun en Hungría, fundó la revista *Scientometrics*. Después de su fundación en 1980, Nalimov se convirtió en el único editor consultor (Granovsky, 2001). En este escenario, los métodos cuantitativos de la cienciometría son ampliamente aplicables a la ciencia de la ciencia, demostrando la relación entre ambas disciplinas.

Además Nalimov, entendió a la ciencia como un proceso de información en el que las citas podrían utilizarse como indicadores (Chiorni & Guilarievkii, 2001). Luego en 1966 llegó la edición de 1965 del *Science Citation Index (SCI)* al instituto donde trabajaba Nalimov. Compartiendo la misma percepción de Eugene Garfield sobre la importancia de las citas, inició un seminario informal sobre análisis de citas y como usar el SCI.

Michael J. Moravcsik (1928-1989) en sus trabajos de 1980 y 1985 (citados por Rosa Sancho) (Sancho, 1990) señala que, "la ciencia se puede estudiar bajo tres aspectos: actividad, productividad y progreso científico; similar tricotomía cabe hacerse con los aspectos de calidad, importancia e impacto científico". Además se

cuestiona a la ciencia como un sistema que podía ser administrado y estudiado desde la perspectiva del análisis costo-beneficio, tal y como sucede en los sistemas económicos (Moravcsik, 1980, 1985, 1989) citado por (Sancho, 1990).

Por su parte, Garfield comenzó a utilizar los índices de citas (Garfield, 1955) en los estudios de redes de publicaciones, junto con los modelos matemáticos de Nalimov sobre el desarrollo científico, convirtió a la información científica en la principal base empírica de la cienciometría. Garfield también estableció el método de co-citation conocido por "ciencia cartográfica". Este método permite revelar la disciplina y la estructura conceptual de la ciencia, la estructura de las comunidades científicas, la evolución de sus transformaciones en el tiempo y la identificación de nuevas áreas de investigación emergentes (Garfield, 1983, 2004). Este enfoque estimuló el desarrollo de la cienciometría y nuevas técnicas desde la cienciometría para hacer ciencia de la ciencia.

Las teorías y métodos de la cienciometría son instrumental para la Ciencia de la ciencia (Fortunato et al., 2018). Particularmente, de la cienciometría toma la idea de medir la ciencia considerando: las fuentes de datos a gran escala como es el *Web of Science*, Scopus, PubMed, Google Scholar, Microsoft Academic, U.S. Patent and Trademark Office, y otros; aplica una amplia colección de métodos cuantitativos basados en los modelos matemáticos de Nalimov (Nalimov & Mulchenko, 1969), Price (Price, 1951), Garfield (Garfield, 1955); también utiliza los modelos de visualización de datos basados en la ciencia de redes (Börner et al., 2020; Börner, Simpson, Bueckle, & Goldstone, 2018; Börner, Theriault, & Boyack, 2005; Chen, 2004) y algoritmos basados en inteligencia artificial (Villaseñor et al., 2017).

### 1.1.2. La Ciencia como sistema complejo: retos de medición y análisis

Los antiguos filósofos griegos introdujeron la ciencia basada en la razón, la evidencia y la ontología, pudiéndose verificar a través de experimentos. En este escenario, el concepto de "sistema" fue muy importante. Su significado evolucionó con el paso del tiempo. René Descartes, Charles Robert Darwin, Nicolas Carnot, Robert Brown, Rudolf Clausius, Henri Poincaré, Albert Einstein, y otros grandes científicos agregaron un valor esencial a nuestra comprensión contemporánea del concepto de sistema.

Después de más de dos mil años, el concepto de sistema sigue siendo una de las herramientas intelectuales clave en el desarrollo de nuestra civilización. Hoy en día, muchos científicos destacados creen que la comprensión del concepto de sistema complejo constituye el desafío más crítico para el desarrollo de la ciencia. ¿Qué se entiende actualmente por sistemas complejo?, ¿Por qué la ciencia es un sistema complejo? y ¿Qué retos enfrentamos como cienciométricos para su medición y análisis? En esta sección se ofrecen posibles respuestas a las interrogantes planteadas.

Muchos son los intentos de formalizar la definición de un sistema complejo, a continuación se muestran algunos:

Whitesides y Ismagilov entendieron que un sistema complejo tiene una evolución muy sensible a las condiciones iniciales o a pequeñas perturbaciones, requiriendo

para su análisis ecuaciones diferenciales no lineales (Whitesides & Ismagilov, 1999).

“En un sistema complejo hay múltiples interacciones entre los diferentes componentes que lo integran ...” (Ladyman, Lambert, & Wiesner, 2013).

Los sistemas complejos son esencialmente una manifestación de leyes de poder debido al comportamiento de sistemas acoplados, distribuidos y no lineales. (Fieguth, 2017).

La autora entiende que, un sistema complejo está compuesto por partes interconectadas que, en su conjunto exhiben una o más propiedades. Además, ellos cumplen la propiedad de ser dinámicos, cambiando con el tiempo y los estados anteriores del sistema pueden influir en los estados actuales.

Una vez planteado el concepto de sistemas complejos, pudiéramos analizar la ciencia desde la perspectiva de los sistemas complejos y además clasificarla como un sistema complejo (Fieguth, 2017; Ladyman et al., 2013). La ciencia misma exhibe una naturaleza compleja, por ejemplo: cuando observamos las conexiones entre los científicos que hacen la ciencia. Autores (Anderson, 2008; Long, 2010) coinciden que la ciencia en su comportamiento sigue los principios de la complejidad de un sistema complejo abierto y no determinista.

Particularmente, la ciencia puede observarse como un sistema complejo que exhibe alta interconectividad entre los elementos que la componen; sus partes son dependientes al cambio de sus partes y se ven afectadas por dicho cambio. Los problemas que se dan en la ciencia tienen múltiples posibles soluciones, es decir, un punto de bifurcación conduce a múltiples posibilidades y por tanto, son disímiles las opciones de respuesta del sistema. El estado actual de la ciencia puede describirse en términos de su historia, es decir, los eventos que ocurrieron en el pasado (Cilliers, 1998; Mitleton-Kelly, 2003).

La ciencia vista como sistema complejo se ha transformado. Los conceptos, herramientas y métodos para estudiarla cada vez son más sofisticados. Estos desarrollos están conduciendo a los científicos sociales (cienciométricos, bibliotecarios, profesionales de la información) en avanzar hacia la formulación de nuevos modelos cuantitativos (Cao, Chen, & Liu, 2016; Deliverable, Vertesy, & Tarantola, 2012; Moed, 2010; Šubelj, Van Eck, & Waltman, 2016) y principalmente el análisis y visualización de grandes volúmenes de datos para comprender la dinámica de la ciencia.

La tendencia hoy es resolver problemas científicos de carácter social con equipos multidisciplinares de investigadores. Debido a la naturaleza cambiante del conocimiento y la estructura social de la ciencia, las prácticas establecidas en la ciencia están experimentando cambios. Analizar la ciencia utilizando métodos científicos es el núcleo de la ciencimetría. Por tanto, el desafío en este escenario es cómo describir la estructura y evolución de la ciencia, así como, replicar y predecir su estructura y evolución.

En la década de los 60, 70 y 80, los cienciométricos tenían como retos clave: mapear los colegios invisibles (Garfield, 1983); estudiar los vínculos entre productividad y calidad; investigar las prácticas de citación en diferentes campos; evaluación de la investigación y el uso del análisis de citas en la formulación de políticas (Small, 1973) (Braun, Glänzel, & Schubert, 1985) y desarrollaron el factor de impacto (IF) (Garfield & Sher, 1963) como medida para evaluar revistas.

Los años 90 y 2000, trajeron consigo el surgimiento de nuevos retos para abordar la ciencia. Por su parte, Leydesdorff planteó el estudio de la ciencia desde un espacio tridimensional de unidades de análisis: dimensiones sociales (personas, instituciones), institucionales (reglas, financiamiento, métricas, indicadores) y cognitivas (textos, revistas) (Leydesdorff, 1995). Payette, caracterizó la estructura de la dinámica de la ciencia a través del modelo *Agent-based modeling* (Payette, 2012). Barabási y colaboradores utilizaron la teoría de redes (Barabási & Réka, 1999) en el análisis de la coautoría, co-citation y co-ocurrencia de palabras.

Además, el análisis de la evolución de las redes de coautores y citas de artículos (Börner, Maru, & Goldstone, 2004) condujo al planteamiento de nuevas técnicas de visualización (Börner, Chen, & Boyack, 2003; Börner et al., 2020; C. Chen, 2006a; Jacomy, Venturini, Heymann, & Bastian, 2014; Wu, Wang, Wang, Zhang, & Meyer, 2020) y análisis de datos para el estudio de la estructura y dinámica de las disciplinas científicas (Boyack, Börner, & Klavans, 2009). Surgieron importantes proyectos (Börner, 2010; Perez-Angon, 2018; ScimagoLab, 2007) que su principal reto era analizar, visualizar y describir las condiciones bajo las cuales surgen los descubrimientos científicos y las áreas donde estos descubrimientos se pueden encontrar. Basándose en modelos matemáticos y computacionales.

Al mismo tiempo, uno de los problemas en este escenario es lo difícil de determinar nombres, variantes de nombres y errores ortográficos para autores e instituciones. Algunas técnicas automatizadas han resuelto parcialmente este problema, pero todavía existen muchas iniciativas encaminadas a ofrecer solución.

El diseño de visualizaciones estáticas para explicar los cambios en la ciencia es solo un punto de partida. Expertos en el tema han reconocido la necesidad de visualizar la dinámica de la ciencia (Boyack, Klavans, & Börner, 2005; Boyack, 2009; Boyack et al., 2009; Yw Chen, Fang, & Börner, 2011; Leydesdorff, Carley, & Rafols, 2013) con el objetivo de ayudar a refutar o confirmar teorías existentes, así como, generar interrogantes que conducen a nuevas investigaciones.

## **1.2. Descubrimiento de conocimiento en bases de datos (KDD): la perspectiva cuantitativa**

En la era moderna, la mayoría de la información está disponible a través de documentos en lenguaje natural no estructurados, esto se debe al crecimiento exponencial de la web, las bibliotecas digitales, la documentación en formato digital, etc. Sin embargo, este gran cúmulo de datos lleva intrínsecamente información estratégica, a la cual no es fácil de acceder a través de técnicas clásicas de recuperación de información.

El descubrimiento de conocimiento en bases de datos o en inglés *Knowledge Discovery in Databases* (KDD) fue acuñado en el primer “workshop” de KDD en 1989 y se define como: el conjunto de acciones que se desarrollan para analizar las grandes cantidades de datos almacenadas por las organizaciones, con el fin de obtener información de interés para mejorar el funcionamiento y rendimiento de las mismas (Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996).

El objetivo principal del KDD es la extracción del conocimiento que se encuentra dentro de grandes cantidades de datos en bruto, identificar los patrones más importantes y significativos, así como, representar los objetivos del usuario. El descubrimiento de conocimiento oculto, es posible gracias a la aplicación de

sofisticadas técnicas como la inteligencia artificial, la minería de datos y la visualización de información, las cuales permiten encontrar patrones y relaciones dentro de los datos permitiendo la creación de modelos, y representaciones abstractas de la realidad.

El valor real de los datos reside en la información que se puede extraer de ellos, información que ayude a tomar decisiones o mejorar la comprensión de los fenómenos que se presentan en la realidad. Hoy, más que nunca, los métodos analíticos avanzados son el arma secreta de muchos negocios exitosos. Empleando métodos analíticos avanzados para la explotación de datos, los negocios incrementan sus ganancias, maximizan la eficiencia operativa, reducen costos y mejoran la satisfacción del cliente (Bath, 2004; Gamarra, Guerrero, & Montero, 2016). El reto no es la extracción de información significativa de los datos no estructurados sino hacer los datos útiles al usuario final.

KDD es una metodología interactiva e iterativa que incluye varios pasos, una de las primeras propuestas fue planteada por Fayyad en 1996 (Figura 1) y en ella el proceso se divide en las siguientes fases: comprensión del problema, selección de la base de datos objetivo del proceso de descubrimiento, limpieza y preprocesado de dicha base de datos, reducción, elección del método que se va a usar para resolver el problema, elección del algoritmo, ejecución del algoritmo de Minería de datos elegido, interpretación de los patrones obtenidos y consolidación del conocimiento obtenido (Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

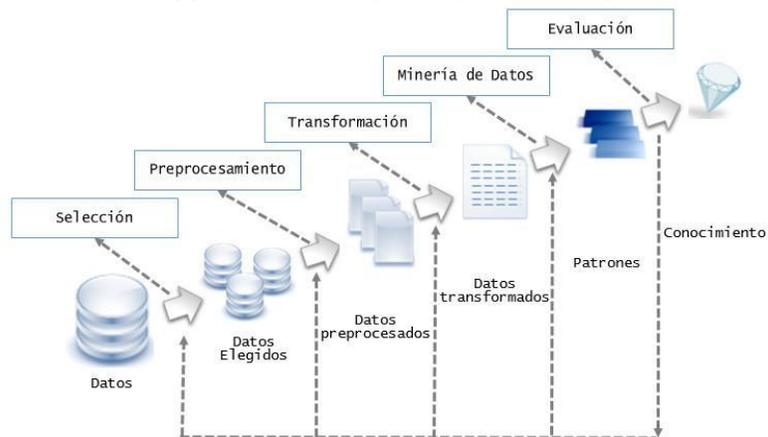


Figura 1. Fases de la metodología para el descubrimiento de conocimiento en bases de datos (KDD)(Fayyad, Piatetsky-Shapiro, & Smyth, 1996).

La metodología KDD puede aplicarse para minar datos bibliográficos en bases de datos como: Scopus, *Web of Science*, MEDLINE. A finales de la década de los años 1980, KDD se introdujo por primera vez en la búsqueda de información en un estudio sobre el síndrome de Raynaud (Swanson, 1986). Posteriormente, esta metodología ha sido aplicada en campos de la ciencia de la información. En la búsqueda de información bibliográfica (Gordon & Lindsay, 1996; Lindsay & Gordon, 1999); en bibliotecas y centros de información para el procesamiento de datos (Marcum, 2001; Nicholson, 2002), en el control de calidad y verificación de la integridad de los datos, lo que se entiende como *online analytical processing* (Hudomalj & Vidmar, 2003).

Otra perspectiva del KDD es el uso de esta metodología en el campo de la cienciometría. Particularmente, con el objetivo de extraer, analizar y visualizar patrones en bases de datos de información científica. El desarrollo de ViBlioSOM, metodología que conjuga los pasos del KDD para el análisis métrico de la información y la visualización de sus resultados en forma de mapas auto-organizados (Sotolongo Aguilar et al., 2002; Villaseñor et al., 2017). Otros estudios (Björneborn & Ingwersen, 2001; He, 1999) han aplicado la metodología en la webmetría, en el análisis de co-citaciones y en el mapeo de la literatura científica (Nieminen, Pölonen, & Sipola, 2013; Wei, Grubestic, & Wade Bishop, 2015).

### 1.2.1. Mapeo científico y visualización de información

La capacidad de visualizar los datos, es tan antigua como la humanidad misma. Los orígenes de la visualización de la información, se identifican por la búsqueda del *insight* a partir de la entropía que existe en un gran volumen de datos, a partir de dos perspectivas: estructura y cambio. Sus antecedentes se localizan en la historia del desarrollo del conocimiento y el lenguaje visual, estrechamente fusionado a la influencia de los progresivos avances tecnológicos y científicos, con dos propósitos esenciales: descubrir y explicar (Chen, 2004, 2017).

El término visualización de la información se manejó por primera vez en 1989. En esta etapa el término refiere a una propuesta de arquitectura cognitiva para interfaces de usuario interactivas. Luego, bajo la influencia de los paradigmas cognitivo-tecnológico y *Human Computer Interaction* (HCI), se fortaleció como área de conocimiento. Consolidándose como un área de investigación, con bases interdisciplinarias y enfocadas en la representación visual de un gran volumen de información; en el uso de técnicas gráficas interactivas para comprender y analizar los datos de forma intuitiva; detectando y explorando sus relaciones y patrones de comportamiento.

La visualización de información es el proceso de interiorización del conocimiento mediante la percepción de la información. También se puede enfocar sobre los aspectos cognitivos y perceptuales de los seres humanos. La informática en este caso es el medio para obtener las visualizaciones y “traducir” lo cognitivo a un lenguaje informático (Boyack, Wylie, & Davidson, 2002).

Chaomei Chen, define el término de visualización de información, como aquellas representaciones gráficas de información procedentes de una computadora que tienen el objetivo de constituir visualmente espacios y estructuras de información que faciliten una rápida comprensión, y la posibilidad de identificar y extraer patrones a partir de una gran cantidad de información (Chen & Song, 2017).

La comunidad científica especializada en visualización de la información ha desarrollado una gran diversidad de especificaciones de software. Estos avances facilitan el análisis de conjuntos de datos a gran escala. Métodos que, comenzaron por la visualización y análisis de co-ocurrencia de palabras para identificar temas semánticos. Luego estas técnicas de visualización han agregado la capacidad de visualizar dominios de conocimiento. Específicamente, en los últimos 20 años hemos visto una explosión cámbrica de mapas científicos.

En este escenario, *Science mapping* o mapeo científico es un proceso genérico del análisis y visualización de dominios conocimiento científico (Chen, 2017). El

mapeo científico es una representación espacial de cómo las disciplinas, campos, especialidades, documentos y autores están relacionados entre sí (Small, 1997).

Katy Börner, define el mapeo científico como el conjunto de mapas que intenta encontrar representaciones de las conexiones intelectuales dentro de un sistema de conocimiento científico cambiante (Börner, Bueckle, & Ginda, 2019). En otras palabras, el mapeo científico tiene como objetivo mostrar los aspectos estructurales y dinámicos de un dominio científico.

Un campo de investigación o dominio de conocimiento puede caracterizarse por su base intelectual y los frentes de investigación; en inglés se denomina *intellectual base* y *research fronts* (Comins & Leydesdorff, 2016; P. Liu, Chen, Liu, & Xie, 2016). La base intelectual es la colección de trabajos académicos que han sido citados por la comunidad investigadora correspondiente, mientras que los frentes de investigación son los trabajos que se inspiran en los de la base intelectual. El mapeo científico facilita la visualización de los cambios que pueden experimentar la base intelectual e identificar frentes de investigación.

Para realizar mapeos bibliométricos, los científicos sociales deben desarrollar habilidades y conocimientos. Habilidades en el manejo de bases de datos *Web of Science*, *Scopus*, *Google Scholar* y *PubMed*. Conocimiento sobre métodos cuantitativos que incluyen análisis de co-citaciones de autor (ACA) (Howard D. White & McCain, 1998), análisis de co-citación de documentos (DCA) (H. Small, 1973), análisis de co-ocurrencia de términos (Callon, Courtial, Turner, & Bauin, 1983) y sobre los indicadores bibliométricos y cuantitativos.

Por su parte, Cobo y colaboradores, plantean que el mapeo científico de un dominio tiene los siguientes pasos: recuperación de datos, preprocesamiento, extracción de redes, normalización, mapeo, análisis y visualización. Al final de este proceso, el analista debe interpretar y obtener algunas conclusiones de los resultados (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2012). A continuación se detallan cada una de las etapas:

*Recuperación de datos:* fuentes bibliométricas donde se pueden recuperar los datos, como ISI *Web of Science* (WoS), Scopus o otras bases de datos. Además, se pueden utilizar datos de patentes.

*Preprocesamiento:* es el paso de mayor importancia porque la calidad de los resultados dependerá de la calidad de los datos. Se aplican métodos para detectar elementos duplicados y mal escritos.

*Extracción de redes:* a través de las técnicas de análisis bibliométrico se pueden obtener las relaciones entre las unidades de análisis: co-ocurrencia de términos, análisis de co-autoría, acoplamiento bibliográfico y co-citación de documentos o autores.

*Normalización:* después de obtener la red de relaciones entre las unidades de análisis seleccionadas se aplica una transformación a los datos para derivar similitudes de los datos. En este caso las medidas más usadas son el coseno de Salton (Salton & McGill, 1983), el índice de Jaccard (Peters & van Raan, 1993a,

1993b), el índice de equivalencia (Callon, Courtial, & Laville, 1991) y la fuerza de asociación (Nées Jan Van Eck & Waltman, 2009; Waltman & van Eck, 2007).

*Mapeo*: etapa para construir el mapa aplicando un algoritmo de mapeo a toda la red. Los algoritmos de mapeo más recientes que se pueden aplicar a la red son: análisis de componentes principales, *clustering* (Chen, Ibekwe-Sanjuan, & Hou, 2010), *Pathfinder* (Chen, 1999) y análisis de redes sociales.

*Análisis*: para extraer conocimiento del mapa construido se pueden aplicar el análisis de redes (Barabási et al., 2002); temporal o *temporal analysis* (Garfield, 2004), *burst detection* (Chen, 2006a) y análisis geoespacial (Cobo et al., 2012).

*Visualización*: la técnica de visualización empleada es muy importante para una buena comprensión y una mejor interpretación del resultado. Mapas heliocéntricos (Moya-Anegón et al., 2005), modelos geométrico (Skupin, Biberstine, & Börner, 2013) y redes temáticas (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011) son las técnicas de visualización más usadas en el mapeo científico.

Diversas son las herramientas de software que actualmente se utilizan en la etapa de visualización. Entre las herramientas más usadas en el mapeo científico podemos mencionar: Bibexcel (Persson, O., Danell, & Wiborg, 2009), CiteSpace 5.7 (C. Chen & Song, 2019), BiblioMaps (Grauwin & Sperano, 2018), bibliometrix R (Aria & Cuccurullo, 2017), el paquete de Softwares de Leydesdorff, Sci2Tool (K. Börner, 2011) y VOSViewer (Nees Jan van Eck & Waltman, 2010). En consecuencia con esta explosión de herramientas, pensamos que, en el mapeo bibliométrico se pueden aplicar varios de estos software para generar otras perspectivas. Por tal motivo, en nuestra metodología para analizar y visualizar perfiles cuantitativos en la ciencia consideramos la perspectiva de alguna de estas herramientas.

## 1.2.2. Técnicas de mapeo científico

Las unidades de análisis más comunes en el mapeo científico son: revistas, documentos, autores y términos o palabras clave. Entre las unidades de análisis se pueden establecer relaciones. Por ejemplo: el análisis de coautores o coautorías (White & McCain, 1998) se puede analizar la estructura social de un campo científico. El análisis de co-ocurrencia de palabras (Callon et al., 1991) se utiliza para mostrar la estructura conceptual y los principales conceptos tratados por un campo. La co-citación (Small, 1973) y el acoplamiento bibliográfico (Kessler, 1963) se utilizan para analizar la estructura intelectual de un campo de investigación científica. En este acápite se describen las técnicas de análisis bibliométrico que se utilizan en el mapeo científico.

### 1.2.2.1. Análisis de citas

Los análisis de citas han estado marcados por la explosión de nuevas bases de datos con algoritmos computacionales, herramientas de software y la disponibilidad de las citas en formato digital. En los últimos años, se ha observado un creciente interés en el análisis de citas para ayudar a abordar diversos temas de investigación: gestión de información, evaluación de la investigación y visualización de dominios

de conocimiento. Considerando que, en la literatura se ha discutido ampliamente las limitaciones del análisis de citas en el estudio de la ciencia y la comunicación académica (Garfield, 1979; Garfield, 1955; White, 1990).

La lista de referencias en un artículo de investigación constituye una parte esencial en el artículo. Las referencias vinculan el artículo actual con publicaciones anteriores y, por tanto, conectan el artículo actual con la red global de publicaciones científicas. Generalmente se asume que una cita representa la influencia y relación (por ejemplo, tema o enfoque metodológico similar) del trabajo citado en el nuevo trabajo de un autor (Leydesdorff, 1998).

El análisis de citas se ocupa del estudio de estos usos y relaciones. Aunque los usos y relaciones individuales pueden ser útiles de examinar, el análisis de citas proporciona principalmente perspectivas macro mediante el uso de grandes conjuntos de datos, explotando el consenso entre un gran número de autores que citan con respecto a la influencia y las relaciones entre académicos y trabajos académicos.

En general, el análisis de citas puede ayudar eficazmente en el descubrimiento de nuevo conocimiento, en la gestión y uso de los recursos de conocimientos existentes (Boyack & Klavans, 2010; Garfield, 1979; Small, 1998). En particular, el análisis de citas se puede utilizar para:

- evaluar los recursos de información y las contribuciones académicas
- mapeo de dominios de investigación para estudiar sus estructuras intelectuales
- seguimiento de los flujos de conocimiento y la difusión de ideas
- estudiar los usuarios y usos de la literatura académica
- ayudar en la organización, representación y recuperación de la información

En el mapeo científico se utiliza el análisis de citas para facilitar la organización del conocimiento. También permite a los investigadores examinar las características, estructuras y evolución de los campos de investigación y las comunidades académicas.

Expertos en estos últimos veinte años han demostrado a través de sus estudios las perspectivas de la ciencia que se pueden analizar aplicando el análisis de citas. Se puede estudiar las estructuras y características de una especialidad y comunidad (Leydesdorff, 1998). Se puede modelar el desarrollo histórico de la especialidad y la comunidad (Bornmann & Leydesdorff, 2015; Egghe & Rousseau, 2002; Hernández-Alvarez, Gomez Soriano, & Martínez-Barco, 2017; Skupin et al., 2013).

Egghe y Rousseau opinan que los análisis de citas se clasifican según tres aplicaciones principales: "evaluación cualitativa y cuantitativa de científicos, publicaciones e instituciones científicas; modelación del desarrollo histórico de la ciencia y la tecnología; y búsqueda y recuperación de información" (Egghe & Rousseau, 1990).

Existen tres tipos de medidas basadas en las citas. Los recuentos entre citas que, es el número de veces que dos documentos se han citado entre sí. Los recuentos de co-citaciones que es el número de documentos que han citado dos objetos juntos. Por último, las frecuencias de acoplamiento bibliográfico, es decir, el número de referencias citadas que tienen dos objetos en común.

La Web ha tenido un gran impacto en los análisis de citas. Con el cambio de siglo, han aparecido decenas de bases de datos como *Scopus*, *arXiv e-print server*, *CiteSeer*, *Physical Review Online Archive* y *Google Scholar*, que calculan los recuentos de citas con una rapidez y facilidad sin precedentes. Esto ha marcado el comienzo de una fuerte competencia para *Clarivate Analytics* y la *Web of Science*. Base de datos que desde 1963 ha tenido el monopolio en los análisis de citas. En las siguientes secciones abordaremos con detenimiento el acoplamiento bibliográfico, la co-citación y la co-ocurrencia.

#### 1.2.2.2. Acoplamiento bibliográfico

El término acoplamiento bibliográfico o *bibliographic coupling* (ACB) fue introducido por el Dr. Kessler del Instituto Tecnológico de Massachusetts (Kessler, 1963). El acoplamiento bibliográfico ocurre cuando dos artículos utilizan una referencia como una unidad de apareo entre ambos. La importancia de esta técnica depende del número de referencias que ambos artículos tienen en común. Los sistemas de software (Aria & Cuccurullo, 2017; Bastian, Heymann, S., & Acomy, 2010; Chen, 2006a; Jacomy et al., 2014; Persson, O. et al., 2009) facilitan el procesamiento automático para identificar y calcular la similitud entre pares de artículos.

ACB se introdujo principalmente con el propósito de recuperar información (Kessler, 1963). Sin embargo, en los últimos años se ha aplicado en la evaluación de la investigación, en el análisis de redes de conocimiento y estudios longitudinales. Particularmente, los estudios de análisis de redes de conocimiento basados en ACB se han realizado a todos los niveles: país, institución, documento, revista, campo de investigación y autor (Boyack & Klavans, 2010; García-Lillo, Claver, Marco-Lajara, Seva-Larrosa, & Ruiz-Fernández, 2020; Nájera-Sánchez, Ortiz-de-Urbina-Criado, & Mora-Valentín, 2020). En los estudios longitudinales, el acoplamiento bibliográfico puede agrupar artículos muy recientes, facilitando la identificación de los frentes de investigación (Persson, 1994).

Esta técnica se utiliza en el mapeo científico para visualizar la relación entre documentos o autores. Además, cuando se aplica el acoplamiento bibliográfico con el análisis de co-citación pueden revelar una imagen más completa del estado actual de la investigación en un campo de investigación (Small, 1997). De esta forma, facilitan el estudio de la trayectoria de desarrollo en la que se encuentra un campo de investigación.

Por su parte, Zhao y Strotmann, propusieron combinar el análisis de acoplamiento bibliográfico de autores o *author bibliographic-coupling analysis* y el análisis de co-citación de autor (ACA). El principal objetivo de esta propuesta era mapear la estructura de las influencias intelectuales internas y externas en la investigación de la ciencias de la información (Zhao & Strotmann, 2008). En el siguiente acápite, se aborda el análisis de co-citación como técnica que examina las influencias intelectuales en un dominio o la base de conocimiento del campo.

#### 1.2.2.3. Análisis de co-citación

El concepto de análisis de co-citación fue definido por Henry Small (Small, 1973) en el Instituto de Información Científica (ISI) de Filadelfia, y paralelamente la investigadora Irina Marshakova, del Instituto de Información Científica Técnica

(VINITI) de la entonces Unión Soviética (Marshakova, 1973). El análisis de co-citación se puede realizar a varios niveles: documento introducido por Small (Small, 1973); análisis de co-citación de autor por White y Griffith (White & Griffith, 1981), y de revistas (McCain, 1991).

Dos documentos se consideran co-citados cuando aparecen en la misma lista de referencias de un tercer documento. El número de documentos cuyas listas de referencias contienen ambos objetos es el recuento de citas conjuntas de estos dos documentos. Esto supone que, cualquiera de sus unidades de análisis (documentos, autores, revistas, categoría temática) están relacionadas con el tema o enfoque metodológico.

White y McCain demostraron que, el análisis de co-citaciones puede ser mapeado con el objetivo de mostrar la cohesión y cambios en la estructura intelectual de un dominio. Específicamente, utilizaron la metodología del análisis de co-citación de autor (ACA). Este análisis identificó los académicos que influyeron significativamente en las ciencias de la información, caracterizó la estructura intelectual y el desarrollo de esta durante tres períodos de 8 años (White & McCain, 1998).

El mapeo de las citas entre revistas (análisis de co-citación de revistas) permite identificar la estructura de los patrones de publicación y comunicación científica en determinados campos científicos. Además, facilita la caracterización de la evolución de la estructura social científica de las investigaciones previas (Osareh, 1996).

Actualmente, los científicos disponen de suficiente capacidad informática (Bastian et al., 2010; Chen, 2006a; Jacomy et al., 2014; Nees Jan van Eck & Waltman, 2010) para analizar y visualizar vastas redes científicas basadas en el análisis de co-citación. Cada sistema de software mapea las relaciones que surgen de los análisis de co-citación. En general, estas herramientas comprenden las fases del proceso de análisis de co-citación.

Frecuentemente, el análisis de co-citación y el acoplamiento bibliográfico se confunden o se consideran erróneamente métodos equivalentes. La diferencia fundamental entre ambos métodos es como se identifican y conforman los grupos de documentos. En el análisis de co-citación, dos o más artículos que comparten uno o más documentos fuente común forman el grupo. Sin embargo, el acoplamiento bibliográfico se centra en grupos de artículos que citan un documento fuente, la co-citación se centra en referencias que vienen frecuentemente en pares (Small, 1998; Small, 1981). Es decir, los procedimientos de co-citación están analizando pares de referencias, es decir, artículos citados, mientras que los métodos de acoplamiento bibliográfico proceden de aquellos artículos que citan y comparten elementos en sus listas de referencia.

En ambas técnicas, los documentos que tienen fuertes vínculos bibliográficos con otros artículos no implica que sus tasas de citas son altas después de su publicación. Los análisis de citas de artículos vinculados bibliográficamente son, por tanto, una importante herramienta de validación que demuestra que los procedimientos de co-citación de citas y acoplamiento conducen a un mayor nivel de exhaustividad en términos de comunicación científica.

#### 1.2.2.4. Análisis de co-ocurrencia de términos

En 1986, Callon, Law y Rip publicaron el libro titulado “*Mapping the Dynamics of Science and Technology*”. La obra definió el concepto de análisis de co-ocurrencia de términos o en inglés *co-word analysis*. Por primera vez esta técnica se plantea para mapear la dinámica de la ciencia empleando la teoría del *actor network* (Callon et al., 1983). Después de la publicación de este libro, la comunidad científica internacional comenzó a aplicar el análisis de co-ocurrencia de términos en los dominios de conocimiento (Yilong Chen et al., 2020; Huang et al., 2014; Shao, Hu, Luo, Huo, & Zhao, 2020).

El análisis de co-ocurrencia de términos es una técnica de contenido eficaz para mapear la fuerza de asociación entre elementos de información en datos textuales (Leydesdorff, 1991). Específicamente, el mapeo científico basado en esta técnica demuestra que un campo de investigación puede caracterizarse a través de las palabras clave más importantes (Borner et al., 2003). Términos que se usan además para medir la similitud entre un par de publicaciones. Los mapas bibliométricos derivados de esta técnica constituyen una herramienta de descripción y pronóstico porque visualizan las relaciones y tendencias en los subcampos. Las visualizaciones, gráficos de red, interconectan los conceptos más significativos e ilustran las asociaciones más fuertes entre los descriptores.

La técnica ha sido aplicada con la intención de estudiar y evaluar los desarrollos, relaciones y la calidad de la investigación en los dominios. En los años 90 y principios de los años 2000 el propósito mayormente era visualizar la estructura de conocimiento aplicando métodos reducción de datos a la matrix de co-ocurrencia (Ding, Chowdhury, & Foo, 2001; Peters & van Raan, 1993a, 1993b). En los últimos años, se combinan métodos para obtener mejor visualización y análisis. Particularmente, la técnica de agrupamiento o *clustering* aplicada a la matriz de datos de ocurrencia de palabras clave y el escalamiento multidimensional de los grupos de palabras resultantes constituyen los avances en este campo (Bornmann, Haunschild, & Hug, 2018; Yilong Chen et al., 2020; de la Hoz-Correa, Muñoz-Leiva, & Bakucz, 2018; Vílchez-Román & Quiliano-Terrerros, 2018).

La tabla 1 muestra un resumen de las técnicas de análisis basadas en las citas, las unidades de análisis/objetos de análisis y el tipo de relación que es posible visualizar en los mapeos científicos:

<b>Técnica bibliométrica</b>	<b>Unidad u Objeto de Análisis</b>	<b>Tipo de Relación</b>	<b>Tipo de mapeo científico</b>
Análisis de citas	Autor Documento Revista País Institución	Co-ocurrencia de Autores Co-ocurrencia de Documentos Co-ocurrencia de Revistas Co-ocurrencia de Países Co-ocurrencia de Instituciones	Red de Co-ocurrencia de Autores Red de Co-ocurrencia de Documentos Red de Co-ocurrencia de Revistas

			Red de Co-ocurrencia de Países Red de Co-ocurrencia de Instituciones
Acoplamiento Bibliográfico (AB)	Referencia del Autor  Referencia del Documento  Referencia del Revista	Referencias comunes entre las obras del autor  Referencias comunes entre documentos  Referencias comunes entre las obras de la revista	Redes de Acoplamiento Bibliográfico de Autores  Redes de Acoplamiento Bibliográfico de documentos  Redes de Acoplamiento Bibliográfico de revistas
Co-citación	Referencia del Autor Referencia del Documento Referencia de la Revista	Co-citación de autor Co-citación de documentos Co-citación de revistas	Red de Co-citación de autor Red de Co-citación de documentos Red de Co-citación de revistas
Co-ocurrencia de términos	Palabras Clave o Términos extraídos del título, resumen, o cuerpo del artículo	Co-ocurrencia de términos	Red de Co-ocurrencia de términos

\*Elaboración propia

### 1.2.3. Sistemas de software para la visualización

Las herramientas para la visualización han transformado en los últimos años el campo de los estudios métricos de la información. Probablemente esto se deba a varias causas. El conocimiento escrito se ha vuelto ampliamente disponible en los formatos digitalizados en los últimos años. El desarrollo explosivo de la *World Wide Web* y las bases de datos en línea, proporcionan enormes cantidades de datos para este tipo de estudio. Además, los bibliométricos y cientímetricos han aprovechado el aumento de la potencia informática para analizar y visualizar la ciencia.

Los estudios métricos de la información se pueden dividir en dos grandes campos de estudio: análisis del desempeño y análisis del mapeo científico (Braam, Moed, & Vanraaf, 1991). El análisis del desempeño tiene como objetivo evaluar los actores científicos (investigadores, instituciones, países, etc.) a través de indicadores cientímetricos basados en datos de publicaciones y citas. Por otro lado, el mapeo científico radica en la representación topológica y temporal de la estructura cognitiva

y social de un campo de investigación (Small, Boyack, & Klavans, 2014). En este escenario, las herramientas de mapeo o visualización constituyen uno de los componentes principales en el análisis y visualización de información.

VOSviewer y CitNetExplorer sistemas de software creados en el “Centre for Science and Technology Studies” Holanda, destacan por las capacidades en el mapeo basado en las técnicas bibliométricas (Nees Jan van Eck & Waltman, 2010). Chaomei Chen, editor y fundador de la revista “Information Visualization” creó CiteSpace (Chen, 2006).

Existen investigaciones publicadas con un alto nivel de profundidad sobre las ventajas, inconvenientes, diferencias y usos de mayor importancia entre las herramientas de visualización. Moral-Muñoz y colaboradores (Moral-Muñoz, López-Herrera, Herrera-Viedma, & Cobo, 2019) publicaron una revisión sobre las herramientas de libre acceso para los mapeos bibliométricos. Pan, Yan, Cui y Hua, evaluaron la utilidad, el impacto y difusión de los software más usados en el mapeo científico (CiteSpace, HistCite y VOSviewer) (Pan, Yan, Cui, & Hua, 2018). Otro grupo de expertos españoles (Gutiérrez-Salcedo, Martínez, Moral-Munoz, Herrera-Viedma, & Cobo, 2018) propusieron técnicas y herramientas de software esenciales para analizar el impacto de un campo de investigación y su estructura científica.

Recientemente, Moral-Muñoz y Cobo publicaron una revisión actualizada de las nuevas funciones, técnicas de visualización y mejoras en las herramientas. Además, se comparan las funciones de las librerías disponibles en R y Python para realizar mapeos bibliométricos; las cuales no se habían considerado en otras revisiones (Moral-Muñoz, Herrera-Viedma, Santisteban-Espejo, & Cobo, 2020).

En general las herramientas de visualización y librerías en R y Python que mayormente usan los científicos ofrecen las técnicas de visualización. Mapas heliocéntricos (Moya-Anegón et al., 2005), modelos geométrico (Skupin et al., 2013), redes temáticas (Cobo et al., 2012) y clustering (Kostoff et al., 2007; Villaseñor et al., 2017; Waltman, van Eck, & Noyons, 2010). Además, con la mayoría de los software y librerías se pueden obtener: gráficos de líneas de tendencia; gráficos de barras; nube de palabras; histogramas; *tree maps*; redes; mapas entre otro tipo de gráficos.

Específicamente, esta metodología integra como técnica de visualización: gráficos de líneas de tendencia; gráfico de barras, redes bibliométricas; histogramas y mapas de clustering. Empleando herramientas de visualización de frontera: librerías de R y Python y Self-organizing maps o mapas auto-organizados (SOMs), basada en Redes Neuronales Artificiales (RNA) (Kohonen, 1990).

#### 1.2.3.1. Redes de cocitación y colaboración

El mapeo científico es un enfoque que se basa en la visualización de redes bibliométricas. Una red bibliométrica representa las relaciones entre un conjunto de entidades. Esta red consta de nodos y enlaces (Chen & Yang, 2010). Los nodos pueden ser: publicaciones, revistas, países, investigadores o palabras clave. Los enlaces indican relaciones entre los pares de nodos. Las relaciones comúnmente estudiadas son: relaciones de citas, co-ocurrencia de palabras clave, y coautoría. En el caso de las relaciones entre citas, se puede hacer una distinción adicional entre relaciones de citas directas, relaciones de co-citación y relaciones de acoplamiento bibliográfico.

Particularmente, las redes de co-citaciones se obtienen a partir de matrices de co-citación de las unidades de análisis (autores, documentos, revistas etc.). Generalmente se transforman en matrices de referencia-referencia para obtener relaciones de co-citación (White & Griffith, 1981).

En las redes de colaboración o co-autoría los nodos representan autores, países o instituciones que pueden estar conectados a más de dos autores, si han escrito un artículo junto. Para la cienciometría este tipo de red puede visualizar los hábitos y patrones de investigación académica. Además, permiten comprender las leyes topológicas y dinámicas que rigen las redes complejas (Barabási et al., 2002).

Ambos tipos de redes bibliométricas responden a varios enfoques de visualización: basado en el distancia (*distance-based approach*), en gráficos (*graph-based approach*), en líneas de tiempo (*timeline-based approach*) (Nées Jan Van Eck & Waltman, 2014), visualizaciones circulares (Börner et al., 2012) y mapas auto-organizados (Skupin et al., 2013; Villaseñor et al., 2017) por solo mencionar algunas.

Particularmente, el enfoque basado en la distancia y en líneas de tiempo predominan en las redes de co-citación y colaboración que se obtienen en esta investigación. En el enfoque basado en la distancia, cuanto menor es la distancia entre dos nodos, mayor es su relación. La técnica más utilizada para determinar la ubicación de los nodos en una visualización basada en la distancia es el escalado multidimensional. Las redes bibliométricas basadas en el enfoque de líneas de tiempo ubican los nodos en un punto específico en el tiempo (Nées Jan Van Eck & Waltman, 2014).

Las redes de co-citaciones y coautoría representan las dimensiones de la ciencia. Sin embargo, en ambos tipos de redes bibliométricas es necesario comprender los conceptos básicos del análisis de redes sociales: vínculos (*ties o links*); densidad (*density*); *path, length, y distance*; por último centralidad (Zhang, 2010). Estas medidas ayudan a identificar los elementos en la red los conectores, expertos, líderes, puentes; donde están los clusters de mayor importancia y quienes lo componen; quién es el núcleo de la red y quienes están en la periferia.

Generalmente, en las redes de co-citaciones, co-ocurrencia y colaboración existen diferencias entre los nodos y la cantidad de vínculos que tienen con otros nodos. En este caso los datos en la matriz de co-citación, co-autoría o co-ocurrencia se deben normalizar. Las medidas de similitud directa (*direct similarity measures*) permiten esta normalización: *association strength* (Nées Jan Van Eck & Waltman, 2009); coseno de Salton (Salton & McGill, 1983); índice de inclusión; índice de Jaccard (Jaccard, 1901). Métricas que se encuentran disponibles en las herramientas de visualización y mapeo científico. En este escenario, es importante distinguir dos problemas: la normalización y el tipo de matriz (simétrica y asimétrica) (Leydesdorff, 2008).

Van Eck y Waltman reportan que *association strength* es la mejor métrica para normalizar las matrices asimétricas de co-ocurrencia. La fuerza de asociación o *association strength* como medida de similitud probabilística cumple con la propiedad de corregir el efecto del tamaño en el número de co-ocurrencias de dos objetos (Nées Jan Van Eck & Waltman, 2009). Luukkonen y colaboradores argumentaron que, el índice de Jaccard es preferible para normalizar datos en una matriz simétrica de co-autoría o colaboración (Luukkonen, Tijssen, Persson, &

Sivertsen, 1993). Por su parte, Leydesdorff, sugiere aplicar el coseno de Salton para normalizar una matriz asimétrica de co-citaciones (Leydesdorff, 2008).

Chaomei Chen, editor de la revista "*Information Visualization*" creó en el año 2006 la herramienta de mapeo científico, *CiteSpace*. Aplicación basada en Java y disponible gratuitamente. Su objetivo principal es el análisis y visualización de las tendencias emergentes y cambios en la literatura científica. El software admite varios tipos de estudios bibliométricos: el análisis de redes de colaboración, de co-palabras, de co-citaciones de autores, de co-citaciones de documentos, el acoplamiento bibliográfico, las visualizaciones de texto y geoespaciales.

Particularmente, para el análisis de co-citación la herramienta realiza un análisis progresivo de redes (*progressive network analysis*) (Chen, 2017). El análisis progresivo de redes se centra particularmente en los nodos que desempeñan un rol crítico en la evolución de una red a lo largo del tiempo. Los nodos críticos son candidatos de puntos de inflexión intelectual.

*CiteSpace* representa la literatura en términos de una red sintetizada a partir de una serie de redes individuales (Chen & Song, 2019). Cada red individual se construye a partir de artículos publicados en un intervalo de tiempo de un año, conocido como *time slide*. *CiteSpace* integra estas redes individuales y forma una visión general de cómo un campo científico ha evolucionando en el tiempo. En nuestra investigación pretendemos ofrecer una visión sobre el estado de las características de la base intelectual (Åström, 2006) y los frentes de investigación (C. Chen, 2006b) en dominios de la ciencia mexicana con una ventaja comparativa mayor a la media mundial.

*CiteSpace* ha sido usado para visualizar la estructura y dinámica de dominios de conocimiento como: cambio climático (Fang, Yin, & Wu, 2018); medicina (Chen, Hu, Liu, & Tseng, 2012), recientemente se aplica este método para la recopilación de datos y búsqueda de conocimiento sobre COVID19 (Chen, 2020) y para identificar la literatura más relevante sobre los sistemas de geoinformación *geographic information systems* (Wei et al., 2015) por solo mencionar algunas aplicaciones.

#### 1.2.4. Inteligencia artificial en estudios cuantitativos

John McCarthy se considera el padre de la inteligencia artificial (IA). Él la define como: la ciencia que desarrolla programas o funciones inteligentes que están asociadas a la inteligencia humana, como el razonamiento, el aprendizaje y la resolución de problemas (Gupta & Dhawan, 2018). La IA ha dominado varios campos, como los juegos, el procesamiento del lenguaje natural, los sistemas expertos, los sistemas de visión, el reconocimiento de voz, el reconocimiento de escritura a mano y los robots inteligentes.

La inteligencia artificial basada en conceptos provenientes de disciplinas como la informática, la biología, la psicología, la lingüística, las matemáticas y la ingeniería ha facilitado la creación de sistemas que comprenden, piensan, aprenden y comportarse como humanos. El componente fundamental en este escenario son las redes neuronales artificiales.

Las redes neuronales artificiales están inspiradas en las redes del cerebro humano. En los últimos años, se han aplicado de manera efectiva en diversos problemas científicos. Particularmente, en la cuantimetría las redes neuronales

artificiales se han aplicado para identificar nuevos tópicos (Özmutlu & Çavdur, 2005); en el mapeo de los cambios en la estructura intelectual de dominios (Ravikumar, Agrahari, & Singh, 2015; Van Den Besselaar & Leydesdorff, 1996); en la predicción de citas (Mistele, Price, & Hossenfelder, 2019) y el estudio más reciente propone el uso de una arquitectura de red neuronal para medir la similitud entre las palabras y párrafos que se encuentran en los artículos (Thijs, 2020).

#### 1.2.4.1. Mapas auto-organizados SOM

Teuvo Kohonen, interesado en comprender la clasificación natural que hace el cerebro, ideó el algoritmo SOM. Se le llama SOM al algoritmo matemático por sus siglas en inglés: *Self-Organizing Maps* y sirve para la generación de mapas auto-organizados. Este algoritmo conduce el proceso de entrenamiento de la red neuronal y produce un mapa topográfico bidimensional para la organización y visualización de información. La distribución topográfica de la información (en este trabajo, información bibliométrica) permite identificar cúmulos o clusters que guardan cierta similitud (Kohonen, 1990).

En términos matemáticos, el SOM define una función del espacio de entrada (información de entrada) a una red de neuronas en el plano (mapa). En otros términos, ésta función define una proyección del conjunto de datos multidimensionales (invisible) a un espacio perceptible visualmente (normalmente bidimensional). La visualización del conjunto de datos permite que las relaciones de similitud que se presentan entre los datos dentro del espacio multidimensional puedan ser observadas en un despliegue bidimensional denominado "mapa auto-organizado" (Kohonen, 2013).

La denominación mapas auto-organizados surge porque el proceso de entrenamiento de la red neuronal y la generación de los mapas no requieren de la intervención del especialista (aprendizaje no supervisado). Este modelo tiene una capa de entrada que recibirá la señal, la dimensión ( $n$ ) dependerá del problema (e.g. cantidad de indicadores bibliométricos utilizados). El dato de entrada será un vector,  $X = (x_1; x_2; \dots; x_n)$ , mientras que la capa de salida (formada por  $M$  neuronas) es la encargada de procesar la información en paralelo y formar el mapa (figura 2). La idea básica del modelo, es crear una imagen de un espacio multidimensional de entrada en un espacio de salida de menor dimensión (reducir dimensiones) (Kangas et al., 1990).

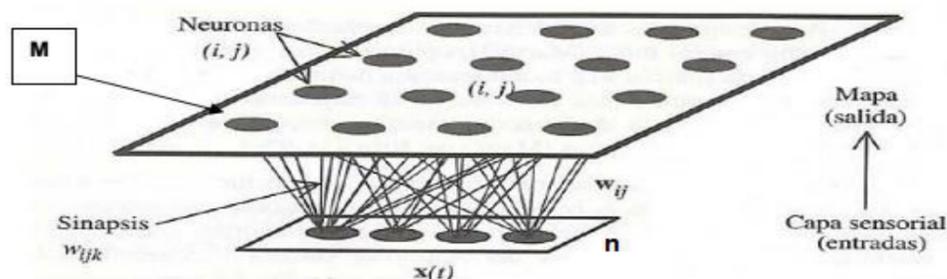


Figura 2. Estructura de una Red Neuronal bajo el algoritmo SOM

Una característica importante en los mapas son las relaciones de semejanza (similarity) o disimilitud (dissimilarity) que se formen entre los datos. Las similitudes que se presentan entre los datos dentro del espacio multidimensional pueden ser observadas en un despliegue bidimensional denominado “mapa”. La semejanza o diferencia se establece cuando se consideran los datos métricos en forma de vectores, por tanto, es posible medir las distancias que se establecen entre ellos a partir de la aplicación de una métrica (fórmula matemática). Se asume que a menor distancia entre dos vectores, existe una mayor similitud entre los respectivos datos métricos (Villaseñor, Carrillo-Calvet, et al., 2017).

#### 1.2.4.2. ViBlioSOM 2.0 – Software

La aplicación de la red neuronal SOM se ha aplicado en el campo de las disciplinas métricas de la información. Los pioneros en emplear este tipo de técnica con fines bibliométricos e incluyendo el análisis de la matriz de palabras clave de co-ocurrencia en los dominios biomédicos fueron Sotolongo, Guzmán y Carrillo. Innovando en esta dirección desarrollaron la metodología ViblioSOM (Visualización Bibliométrica con el Algoritmo SOM), para el análisis métrico y la visualización de la información bibliográfica digital (Sotolongo Aguilar et al., 2002).

Esta metodología consta de una cadena secuencial de procesos para el descubrimiento y la visualización de conocimiento. Este método ocupa la red neuronal SOM para llevar a cabo la minería de datos y la visualización de información, siguiendo la secuencia natural del KDD. Metodología que se ha aplicado en estudios que describen los niveles de granularidad como: en el análisis de patentes (Díaz-Pérez et al., 2014), en la caracterización cuantitativa multidimensional de las instituciones mexicanas más productivas en Scopus (Villaseñor, Carrillo-Calvet, et al., 2017), en la identificación de revistas con un alto impacto (Arencibia-Jorge, Villaseñor, Lozano-Díaz, & Calvet, 2016).

Otros autores han aplicado el algoritmo SOM en los estudios métricos de la información (Boyack et al., 2011; Nieminen et al., 2013; Skupin et al., 2013) y en el mapeo de la ciencia y la tecnología (Guerrero, Moya-Anegón, & Herrero-Solana, 2002; Moya-Anegón, Herrero-Solana, & Jiménez-Contreras, 2006). Estas investigaciones aplican el algoritmo SOM en el procesamiento y visualización de un gran corpus de datos bibliográficos con una metodología y software diferente a los estudios realizados en el Laboratorio de Dinámica no Lineal de la Universidad Nacional Autónoma de México.

Particularmente, el Laboratorio de Dinámica No Lineal de la Facultad de Ciencias de la Universidad Nacional Autónoma de México desarrolló el software ViBlioSOM 2.0 (Jimenez Andrade, Villaseñor-García, & Carrillo-Calvet, 2019). Esta herramienta es de libre acceso, implementada en C # y se ejecuta en el sistema operativo Windows y .Net 4.6 instalado. Con ViBlioSOM 2.0 es posible obtener varios escenarios gráficos y visualizaciones: “Mapa de clústeres”, U-Matrix y “Mapa de componentes”, los cuales se aplican en esta investigación.

Particularmente, el proceso de entrenamiento de la red neuronal en ViBlioSOM 2.0, ocurre tal como se muestra a continuación:

Paso 1	Un vector de entrada es conectado a todas las neuronas en paralelo vía los vectores de referencia de cada neurona: $w_i(t-1) = (x(t)1, \dots, x(t)n)$
Paso 2	Las neuronas compiten para ver cuál de ellas es capaz de representar de mejor manera al dato de entrada: $w_i(t-1)$ vs $(x(t)1, \dots, x(t)n)$
Paso 3	La competencia consiste en encontrar el vector de referencia más cercano al dato $x(t)$ : $\ x - w_c\  = \min_{i=1}^N \{\ x - w_i\ \}$
Paso 4	La neurona ganadora enseña a representar el dato de entrada a sus vecinas: $N_c(t) = \{i \in \mathbb{N} \mid \ r_{c(t)} - r_i\  \leq \rho(t)\}$

Este sistema de software permite aplicar algoritmos de clustering con el objetivo de identificar grupos de datos que comparten características similares. En este trabajo se conoce como clustering jerárquico aglomerativo. Este algoritmo empieza poniendo a cada dato en su propio grupo. Posteriormente, busca los dos grupos más parecidos para unirlos y así crear un nuevo cluster.

Una vez conformado el nuevo grupo, el algoritmo se repite para ir uniendo dos grupos en cada iteración hasta que todos los datos queden agrupados en un solo cluster. En la primera iteración, las comparaciones son entre datos individuales, pero a medida que avanza el algoritmo, se tienen que comparar grupos conformados por dos o más datos.

Existen varias estrategias para definir la distancia entre dos grupos de datos. Dado que la finalidad es unir dos grupos, el término utilizado en la literatura para referirse a estas estrategias es “unión o enlace” (en inglés linkage). Las opciones disponibles son: la distancia promedio entre los datos de un cluster con los datos del otro (average-linkage), la distancia entre los datos más cercanos de los dos grupos (single-linkage), la distancia entre los datos más distantes de los grupos (complete-linkage) y la cuarta opción define la distancia entre dos grupos como el incremento en la suma de las distancias al cuadrado entre los datos al unir los dos clusters (Ward). En este trabajo se evaluaron las cuatro opciones y se optó por la versión Ward.

Cuando el algoritmo de clustering termina, ofrece una colección de agrupamientos (*clusterings*) que van desde el agrupamiento inicial en el que cada campo está en su propio cluster hasta el agrupamiento con todos los datos en un sólo cluster. Para determinar el número óptimo de grupos ViBlioSOM 2.0 calcula varios indicadores: *Silhouette Coefficient*, *Calinski-Harabasz Index* y *Davies-Bouldin Index*. De acuerdo a los experimentos implementados con los datos de esta

investigación decidimos basar nuestra selección del número óptimo de clusters en las puntuaciones obtenidas con el índice Davies-Bouldin.

Las características de esta tecnología permiten el análisis multiparamétrico y al mismo tiempo observar los cambios en el desempeño de las unidades científicas a través del tiempo. Esta idea se fundamenta en el mapeo longitudinal que planteó Garfield (Garfield, 2004).

#### 1.2.5. Perfiles cientiométricos multidimensionales

El análisis cuantitativo de la ciencia se está convirtiendo en *big data*. Con un nivel cada vez mayor de “informatización”, en la que se combinan grandes y heterogéneos conjuntos de datos sobre diversos aspectos de la ciencia, tecnología e innovación. Particularmente, el análisis de la producción científica, como una de las salidas de la ciencia, se está volviendo cada vez más compleja debido a su naturaleza multidimensional. La cientimetría, como estudio cuantitativo de la ciencia, ofrece un enfoque prometedor para comprender perfiles cientiométricos multidimensionales.

En análisis y visualización de datos multidimensionales es un desafío. Este problema ha sido abordado desde diferentes perspectivas. En la descripción del perfil bibliométrico de investigadores altamente influyentes (Palvia, Serenko, Benitez, & Liang, 2020; Peters & van Raan, 1994); en la evaluación de la investigación (Moed & Halevi, 2015); en el análisis del perfil cientiométrico de países (Eito-Brun & Ledesma Rodríguez, 2016; Luwel, 2000; Schulz & Manganote, 2012; Thelwall & Levitt, 2018); en la caracterización de dominios de conocimiento (Daraio et al., 2018; Dwivedi, Garg, & Prasad, 2017; B. Liu, Zhang, & Wang, 2017), en la visualización del perfil científico multidimensional de revistas (Kim, Feng, & Zhu, 2020; Willett & Restrepo, 2017).

En la literatura se han analizado perfiles cientiométrico y bibliométricos multidimensionales aplicando diferentes tecnologías y metodologías: *clustering* de redes (Chen & Song, 2019; Nees Jan van Eck & Waltman, 2017) y otros algoritmos de *clustering* (Hu & Rouseaub, 2009; Ibáñez, Larrañaga, & Bielza, 2013; Pereira, Escuder, & Zanetta, 1998; Thijs & Glänzela, 2009).

En algunos escenarios es deseable obtener una representación visual; especialmente en aquellos casos en los que se desea una combinación de indicadores cientiométricos y bibliométricos. El uso del SOM sirve para crear comparaciones entre unidades científicas que consideren diferentes dimensiones del desempeño científico (Villaseñor et al., 2017).

El concepto de perfil cientiométrico multidimensional de una unidad de análisis (país, institución o campos de investigación) en nuestra metodología está definido por un conjunto de indicadores que miden diferentes dimensiones de la producción científica. Específicamente, el análisis y visualización de la evolución de los perfiles cientiométricos multidimensionales se obtiene con la aplicación de técnicas de *machine learning*.

La metodología diseñada coloca el concepto de perfil de desempeño en el primer plano. Proporcionando los recursos para realizar comparaciones de perfiles y monitorear la evolución del perfil multidimensional.

### 1.3. Indicadores, herramientas bibliométricas y cienciométricas

La industrialización de la ciencia ha revolucionado la disponibilidad de conjuntos de datos que capturan las principales actividades científicas. La información contenida en las publicaciones científicas es necesaria para generar nuevo conocimiento y analizar conjuntos de datos a gran escala con información relativamente más completa. Para realizar este tipo de estudios es necesario el uso de bases de datos que contienen la producción científica.

Un ejemplo representativo es el *Science Citation Index* (SCI) elaborado por el *Institute for Scientific Information* (ISI) y creado por Eugene Garfield en 1957. Los datos del SCI atrajeron a sociólogos de la ciencia por un lado, y un conjunto heterogéneo de científicos de las ciencias naturales principalmente involucrados en “la ciencia de la ciencia o science of science” por el otro. En este escenario, mientras Garfield enfatizaba en el uso del SCI para localizar documentos (SCI como herramienta bibliográfica) el sociólogo de la ciencia Robert Merton, el físico e historiador Derek Price y el cristalógrafo John Desmond Bernal percibieron al SCI como el instrumento esperado para estudiar y orientar la ciencia de una manera objetiva, cuantitativa y desde la perspectiva “científica” (Skalska-Zlat, 2001). El SCI incluye miles de revistas altamente influyentes en las disciplinas y ahora está disponible en línea a través de las plataformas, como Web of Science (WoS).

En 1992 esta institución fue adquirida por *Thomson Scientific & Healthcare*, y luego se conoce como Thomson Reuters, pues la compañía *Reuters* compró *Thomson* en 2008. En el año 2016, *Thomson Reuters* fue vendida a las firmas de capital privado *Onex Corporation* y *Baring Private Equity Asia*; convirtiéndose en *Clarivate Analytics*.

Este recurso ofrece acceso a grandes volúmenes de datos y ha facilitado el estudio de los patrones de producción científica utilizando modelos matemáticos y computacionales. Además, la producción científica es analizada desde diferentes perspectivas calculando los indicadores cienciométricos. Indicadores que actualmente son empleados por académicos, editores de revistas, bibliotecarios y responsables de políticas científica. Por ejemplo, el índice h (Hirsch, 2005) para evaluar científicos y el factor de impacto (Garfield, 1955) para revistas científicas, se utilizan ampliamente en la práctica. A continuación, se describen con más detalle las herramientas que posee este recurso.

*Clarivate Analytics* posee y opera una colección de servicios basados en suscripciones. Esta empresa se enfoca en el análisis de la investigación científica, patentes, estándares regulatorios, protección de marcas registradas, inteligencia farmacéutica y biotecnológica, protección de marcas de dominio y gestión de propiedad intelectual. Los servicios se ofrecen a través de las siguientes plataformas: *Web of Science*, *Cortellis*, *Derwent Innovation*, *Derwent World Patents Index*, *CompuMark*, *MarkMonitor*, *Techstreet*, *Publons*, *EndNote* y *Kopernio*.

La plataforma *Web of Science* tiene un total de 14 base de datos, las cuales están clasificadas en cuatro tipos de colecciones. La colección regional (*Regional Hosted Collection*) contiene bases de datos con contenido regional como: Scielo Citation Index, Russian Citation Index, Chinese Citation Index y KCI Korean Journal Database. Las bases de datos con un enfoque disciplinar como: Medline, BIOSIS Citation Index y Zoological Record se clasifican en la colección temática (Specialty

Collections). Derwent Innovations Index (patentes) y Data Citation Index (conjuntos de datos y estudios de datos) pertenecen a la Colección de datos y patentes (Patent & Data collections).

La colección principal del *Web of Science* (*Web of Science Core Collection*) se compone de:

- Science Citation Index Expanded (SCI-EXPANDED) –1900 presente
- Social Sciences Citation Index (SSCI) --1956-presente
- Arts & Humanities Citation Index (A&HCI) --1975-presente
- Conference Proceedings Citation Index- Science (CPCI-S) –1990 presente
- Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH) --1990-presente
- Book Citation Index– Science (BKCI-S) --2005-presente
- Book Citation Index– Social Sciences & Humanities (BKCI-SSH) --2005-presente

### 1.3.1. Herramientas bibliométricas del *Web of Science*

*Web of Science*, incluye *InCites Benchmarking and Analytics*, *InCites Essential Science Indicators*, e *InCites Journal Citation Reports*, herramientas analíticas que procesan los datos bibliográficos contenidos en la colección principal.

Particularmente, *InCites Benchmarking & Analytics* (*InCites B&A*) es una herramienta de evaluación de la investigación personalizada basada en la web. *InCites B&A* permite analizar la productividad institucional, monitorear la actividad de colaboración, identificar investigadores influyentes, y descubrir campos de investigación con mejor desempeño. *InCites B&A* procesa los datos de siete bases de datos de la colección principal del *Web of Science* desde 1980 hasta la fecha (Figura 3). Específicamente, las citas de ESCI se excluyen de forma predeterminada, pero se pueden incluir seleccionando la casilla de verificación con "ESCI" en cada generador de informes.

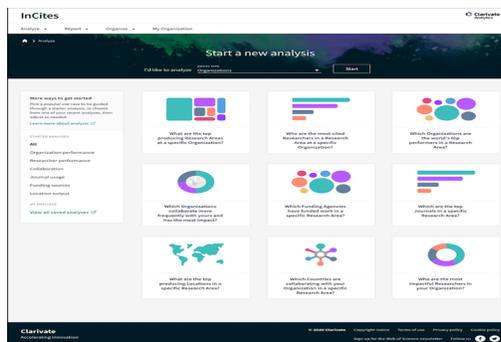


Figura 3. InCites Benchmarking & Analytics (InCites B&A)

Esta herramienta se actualiza bimensualmente y los datos se extraen de la colección principal del WoS. Esto significa que, existe una fecha de corte de datos en el WoS y una fecha de publicación estos datos en el *InCites B&A*. Por ejemplo, la actualización del conjunto de datos en el *InCites* del 12 de agosto de 2018 incluye el contenido de *WoS Core Collection* hasta el 30 de junio de 2018.

*InCites B&A* tiene disponible un total de 17 esquemas de clasificación temática. Tres son exclusivos de *Clarivate* y los otros 14 esquemas pertenecen a otros organismos internacionales. En el acápite 1.3.1.4 se detallan las características de los esquemas de clasificación temática de *Clarivate*, que se consideran en esta investigación. Esta herramienta permite obtener indicadores que son difíciles de calcular, por tal motivo hemos desido incluirla en la metodología.

#### 1.3.1.1. InCites Benchmarking & Analytics: sus clases de métricas

*InCites B&A* permite analizar el desempeño de la investigación combinando indicadores basados en las citas e indicadores cuantitativos. Esta herramienta ofrece los recuentos y promedios simples. Comprende 40 indicadores los cuales están clasificados en: productividad, impacto, colaboración, reputación, acceso abierto, posición del autor y la clasificación de otros para especificar, país, región, estado o provincia, tipo de organización y nivel departamental. La mayor ventaja de *InCites B&A* es el cálculo de los indicadores normalizados y los baselines (rendimiento promedio). Ellos proporcionan una comprensión más precisa y contextual del desempeño de la investigación en estudios comparativos.

Particularmente, para el uso e interpretación correcta de los indicadores es importante considerar varios elementos clave que se mencionan a continuación:

a) La normalización se realiza por tres tipos de contexto (categoría, año y tipo de documento). La plataforma normaliza de esta manera porque las citas entre categorías varían significativamente, crecen con el tiempo y para cada tipo de documento fluntúan las citas. Ejemplo de indicadores normalizados: el Journal Normalized Citation Impact (JNCI) y el Category Normalized Citation Impact (CNCI).

b) El percentil indica como se ha comportado un artículo en relación con otros en su categoría, año y tipo de documento. Los indicadores basados en el percentil ( % documents in top 1%, % documents in top 10% y average percentile) son calculados automáticamente. El sistema crea una distribución de frecuencia por citas recibidas para todos los artículo publicados en el mismo año, categoría y tipo de documento (organizando el artículos en orden descendente según las citas recibidas). De esta manera se identifica el porcentaje de artículos en cada nivel de citación.

c) Baselines o rendimiento promedio se calcula para un conjunto global de artículos en la misma categoría, tipo de documento y año. Este tipo de normalización emplea los conteos completos, es decir, todos los artículos en una categoría temática se consideran independientemente de si esos artículos también están en otras áreas temáticas o no.

La aplicación e interpretación de varias métricas conjuntamente permite obtener una imagen más completa del rendimiento de las citas para un conjunto de unidades

de análisis (autores, instituciones, áreas de investigación, países etc.). Los percentiles se pueden utilizar para complementar otros indicadores como el impacto de las citas normalizado por categoría.

Específicamente, la métrica % de documentos citados (*% Docs Cited*) muestra la eficiencia de un conjunto de documentos. Los indicadores, *% documents in top 1%* y *% documents in top 10%* proporcionan más información sobre la distribución de citas de su conjunto de documentos y nos ayudan a comprender los valores atípicos que se pueden presentar en el *Category Normalized Citation Impact* (CNCI). Por ejemplo, si el valor de CNCI es alto para una universidad, pero tiene un porcentaje bajo de documentos en el top 1% o top 10%, esto sugiere que algunos artículos muy citados son responsables del CNCI alto.

### 1.3.1.2. Essential Science Indicators (ESI)

Es un recurso que forma parte de la plataforma *InCites*. ESI es una herramienta para identificar exclusivamente los artículos de excelencia. Este producto contiene las cifras de las citas para los países, organizaciones, revistas y científicos. Calcula los índices para más de 10,000 revistas en ciencia, tecnología, medicina y ciencias sociales. El análisis va dirigido a dos tipos documentales: artículos científicos y revisiones; se omiten las cartas al editor, erratas, artículos biográficos y otros tipos de documentos (Figura 4).

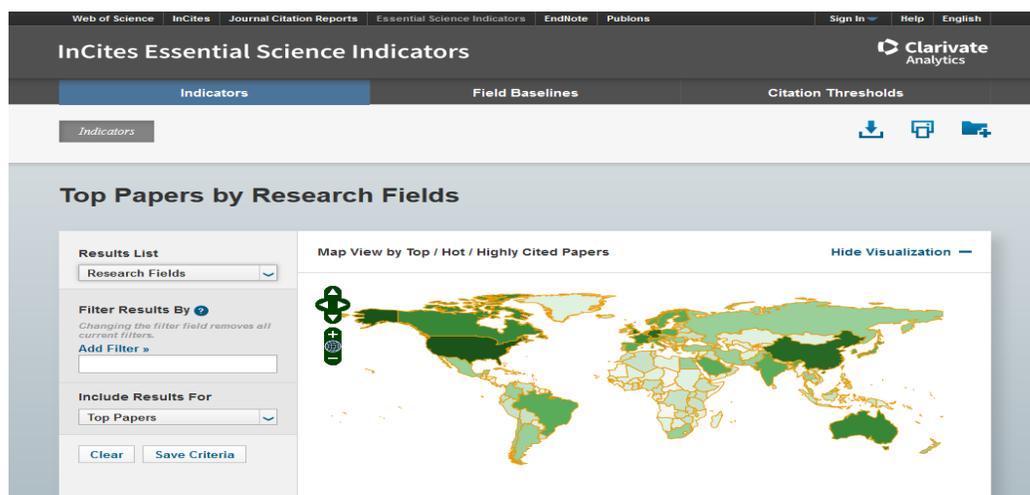


Figura 4. Essential Science Indicators (ESI)

*Essential Science Indicators*, permite realizar complejos análisis de la literatura científica, como por ejemplo: el rendimiento de la investigación de autores, instituciones, naciones y revistas; identificar tendencias significativas en las ciencias y las ciencias sociales. Ofrece el *Top Ranking* de países, autores, instituciones, países/regiones, revistas, 22 campos de investigación y frentes de investigación considerando indicadores de excelencia. En la herramienta existen tres pestañas, cada una destinada a ofrecer tipos de información bibliométrica:

- “*Indicators*”, donde se pueden comparar países, instituciones, autores, revistas, temáticas, campos de investigación y frentes de investigación emergentes de acuerdo con el número de *Top papers*, *Highly cited* y *Hot papers*.

- “*Field Baseline*”, incluye a su vez tres pestañas: (*Citation Rates*) donde se muestra el promedio de citas por artículo y año para cada área de investigación; (*Percentiles*) en ella se definen los niveles de citas para cada una de los 22 campos de investigación; es decir, la cantidad mínima de citas necesarias para que un artículo se encuentre el en 50%, 20%, 10%, 1%, 0.1%, y 0.01% de una misma disciplina y la pestaña “*Field Rankings*”, muestra un ranking de las 22 áreas de investigación en ESI en un período de 10 años y agrega además los conteos del número de *highly cited papers*.

- “*Citation Thresholds*” (Umbral de citas): muestra el umbral de citas, es decir, el número mínimo de citas recibidas, cuando se obtiene un ranking de artículos para cada campo de investigación en orden descendente de citas, luego se selecciona la fracción superior de artículos.

El umbral ESI revela el número de citas recibidas por el top 1% de los autores y de las instituciones; en el caso de los países y las revistas considera los que quedaron en el top 50% superior de una misma disciplina, en un período de 10 años.

### 1.3.1.3. Journal Citation Reports (JCR)

Proporciona una batería de indicadores basados en el impacto de la revista. *Journal Impact Factor* constituye el indicador de mayor importancia para medir la influencia de la revista. JCR cubre revistas indexadas en “*Science Citation Index Expanded*” (SCI) y “*Social Sciences Citation Index*” (SSCI). Esta herramienta no ofrece cifras sobre las citas de las revistas indexadas en “*Arts and Humanities Citation Index*” (A&HCI). Cubre más de 11,500 revistas de 2,550 casa editoriales de 80 países, las cuales pertenecen a 230 ramas del conocimiento (Figura 5).

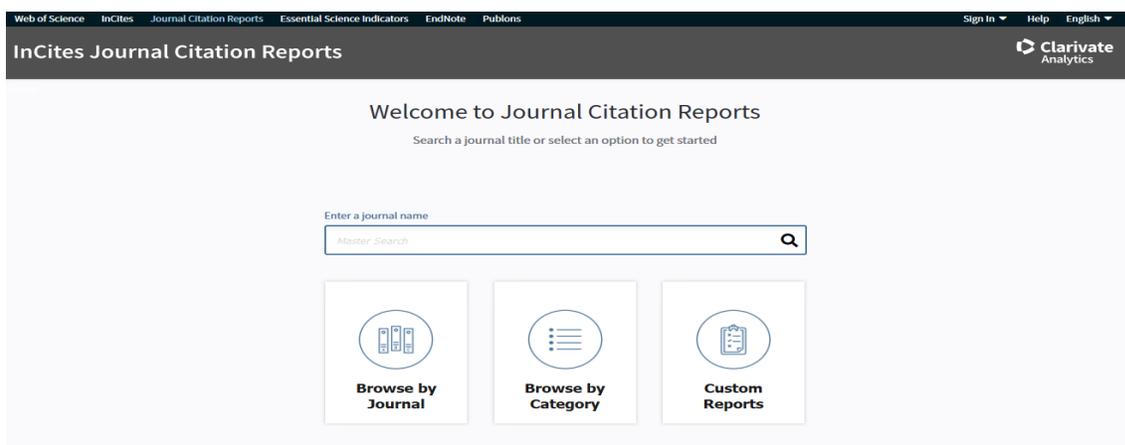


Figura 5. Journal Citation Reports (JCR)

JCR es una herramienta que apoya las tareas de selección y desarrollo de colecciones en las bibliotecas. Los editores la emplean para determinar la influencia de las revistas en una rama de la ciencia y así establecer estrategias editoriales. Los gerentes de investigación y analistas de información para rastrear patrones bibliométricos que sirven para el apoyo y toma de decisiones.

#### 1.3.1.4. Esquemas de clasificación en la plataforma *Web of Science*

Los esquemas de clasificación para las revistas en las áreas de investigación es un tema esencial para los estudios bibliométricos. Los sistemas de clasificación pueden ayudar con varios problemas; por ejemplo, identificar límites en las categorías para evaluar; también para comparar el impacto de la investigación en diversos campos científicos y estudiar la interdisciplinariedad de la investigación.

*Clarivate Analytics*, tiene doce esquemas de clasificación y están disponibles a través de *InCites*. Tres son exclusivos de *InCites* y se describen a continuación:

**Web of Science Core Collection:** El esquema de clasificación del *Web of Science* está compuesto por 252 categorías temáticas (categorías WoS) en ciencias, ciencias sociales, artes y humanidades. Se asigna cada revista a una o más categorías temáticas. Cada artículo publicado heredará todas las categorías de materias asignadas a la revista principal. En este caso el volumen de citas puede variar significativamente de una categoría a otra. Este esquema ofrece mayor nivel de detalle para realizar estudios sobre la productividad científica por dominios de conocimiento.

**Essential Science Indicators (ESI):** comprende 22 áreas de investigación en ciencias y ciencias sociales. Cada revista es asignada a solo una de las 22 áreas temáticas y no hay superposición entre las categorías, lo que facilita análisis más simples. Revistas (como *Science*, *Nature* o Actas de la Academia Nacional de Ciencias) forman la categoría número 22, *Multidisciplinary* (Multidisciplinaria). Los artículos de revistas multidisciplinares son reclasificados en campos específicos a través de un procedimiento automatizado.

Cuando se selecciona el esquema ESI en la herramienta *InCites B&A*, los totales en los resultados y el cálculo de los indicadores basados en las citas solo incluirán los artículos y revisiones del *Science Citation Index Expanded* y el *Social Science Citation Index*. Se excluyen de los resultados las publicaciones de *Arts & Humanities*, *Conference Proceedings Citation Index* y *Book Citation Index*.

**GIIP(Global Institutional Profiles Project) o Institutional Profiles:** Esta clasificación propone una categorización muy amplia. El esquema comprende seis disciplinas y cubre todos los campos de conocimiento. GIIP reclasifica las 252 categorías del WoS a cada una de sus seis disciplinas. Esta agregación contiene una superposición significativa entre disciplinas.

Los otros esquemas se basan en la asignación de datos a sistemas de clasificación de entidades externas. Estos esquemas están diseñados para permitir el uso de indicadores bibliométricos en el contexto de un programa regional de evaluación de investigación, por ejemplo, el Marco de Excelencia en Investigación en el Reino Unido. Alternativamente, el esquema de clasificación de temáticas de la

Organización para la Cooperación y el Desarrollo Económicos (OCDE), herramienta para analizar los indicadores bibliométricos a nivel nacional en el contexto de los datos demográficos y financieros proporcionados por la OCDE. En general, los esquemas basados en clasificaciones de materias externas se desarrollan en asociación con organismos de evaluación de investigación en esa región. Pueden basarse en clasificaciones de revistas o en el mapeo de categorías de *Web of Science*.

Los artículos publicados en revistas multidisciplinarias como *Nature* y *Science* mayormente se centran en un área de investigación. *Clarivate Analytics* reasigna estos artículos a su área temática más relevante. La asignación se realiza algorítmicamente a un área temática al utilizar la información que se encuentra en las referencias citadas de cada publicación. En el caso que los artículos no contengan referencias citadas los artículos se dejan como multidisciplinarios. El proceso de reclasificación permite comparar artículos similares. Este proceso se aplica al área “*Multidisciplinary*” en el esquema *Essential Science Indicators* y en las categorías “*Multidisciplinary Sciences*” y “*Medicine, General and Internal*” del esquema de WoS.

### 1.3.2. Indicadores cuantitativos

La tarea central en la investigación cuantitativa es el desarrollo de métodos y técnicas para el diseño, construcción y aplicación de indicadores cuantitativos sobre aspectos importantes de la ciencia. Particularmente, los indicadores cuantitativos permiten medir el progreso científico en términos de crecimiento y envejecimiento de la literatura. Los indicadores describen el pasado reciente de manera tal que puedan informarnos sobre el futuro cercano. No solo son instrumentos en el estudio de la ciencia sino también tienen la capacidad de describir la interacción entre la ciencia y la sociedad.

Frecuentemente, resulta necesario utilizar estos indicadores para realizar comparaciones entre países, instituciones, campos. Este tipo de comparaciones es un tema delicado. Cada unidad de análisis presenta diferencias en las prácticas de publicación, citación y autoría, lo que dificulta garantizar la equidad de las comparaciones (Waltman & van Eck, 2015). Por ello, la comunidad ha adoptado la práctica de calcular indicadores dependientes e independientes del tamaño, así como, los indicadores normalizados. En esta sección abordaremos las formas en las que estos indicadores son normalizados, así como, otros factores que están presentes en este escenario.

#### 1.3.2.1. Indicadores dependientes e independientes de tamaño

Los indicadores de tamaño dependiente (*size-dependent indicators*), tienen como objetivo proporcionar una medida general del desempeño científico, ejemplo de ello es: el conteo de citas, el número de artículos altamente citados (*highly cited paper*), el índice H, hot paper; y los indicadores independientes del tamaño (*size-independent indicators*) como el promedio de citas por artículo, factor de impacto, tienen por objetivo proporcionar una medida del rendimiento promedio por artículo, revista, institución o país, según sea el objeto de estudio (Waltman, 2016).

Para comparar unidades con niveles de producción y citas diferentes, es necesario calcular indicadores que se clasifican como “independientes del tamaño” (*size-independent indicator*). Este tipo de indicador tiene el objetivo de proporcionar una medida del desempeño promedio por publicación. De esta manera las comparaciones se vuelven justas (Glänzel, Thijs, & Debackere, 2016).

En esta investigación se calculan indicadores de producción e impacto independientes del tamaño y normalizados por las citas, a pesar que, existen argumentos en contra de su uso (Abramo & D’Angelo, 2016). Este tipo de indicadores se aplican correctamente porque miden dimensiones importantes de la productividad y el impacto de la investigación (Glänzel et al., 2016). Estadísticamente son confiables y se basan en modelos matemáticos apropiado.

### 1.3.2.2. Indicadores normalizados

La literatura sobre los indicadores normalizados por campo es extensa. En este caso solo nos enfocaremos en los indicadores de producción e impacto normalizados por campo. La normalización de los indicadores de producción e impacto de las citas comenzó a mediados de los años ochentas (Schubert & Braun, 1986).

La normalización por campo (*field normalization*) en los indicadores cuantitativos tiene como objetivo eliminar diferencias entre campos, por ejemplo, diferencias en las prácticas de publicación, colaboración y citas (Herranz & Ruiz-Castillo, 2012; Waltman & van Eck, 2013).

Particularmente, la mayoría de los indicadores normalizados por campo utilizan un sistema de clasificación para asignar las revistas a uno o varios campos. Ejemplo de ello es: el WoS que ofrece un sistema de clasificación de 252 campos; donde la revista es asignada a uno o varios campos. En cambio en el *Essential Science Indicators* herramienta basada en el WoS el sistema de clasificación contiene 22 campos y cada revista se asigna exclusivamente a un solo campo. Las características de estos sistemas se abordaron con mayor profundidad en la sección 1.3.1.3.

Otra aproximación sobre la normalización por campo en los indicadores es aquella que no requiere de un sistema de clasificación. Este tipo de normalización se conoce en la literatura con varias acepciones: normalización de la publicación citante (*citing-side normalization*) o *fractional citation weighting* (Zitt, 2010), *fractional citation counting* (Zhou & Leydesdorff, 2010), *source normalization* (Moed, 2010), o *priori normalization* (Glänzel, Schubert, Thijs, & Debackere, 2011). La normalización de la publicación citante realiza una corrección en la longitud de la lista de referencias de las publicaciones citantes (Zitt, 2010).

En nuestra investigación los indicadores de impacto normalizado por campo emplean el sistema de clasificación de la base de datos WoS y del ESI. En la siguiente tabla aparecen ejemplos de indicadores que usan y no usan un sistema de clasificación.

Tabla 2. Ejemplo de indicadores normalizados por campos que usa y no usan un sistema de clasificación

Indicadores que usan un sistema de clasificación	Indicadores que no usan un sistema de clasificación
% Documentos Top1 %	Source Normalized Impact per Paper (SNIP)
% Documentos Top10 %	Eigenfactor
Category Normalized Citation Impact (CNCI)	Article influence

Los indicadores de productividad (eficiencia) normalizados por campo realizan una corrección de las diferencias entre campos, considerando la velocidad a la que se producen las publicaciones y se reciben las citas (Waltman & van Eck, 2013). En este caso Abramo y colaboradores propusieron el *fractional scientific strength* (FSS) (Abramo & D'Angelo, 2015). Particularmente, en nuestra investigación se calcula el Índice de productividad nacional (*National Scientific Productivity*) (Villaseñor et al., 2017).

Algunos especialistas cuestionan la normalización por campo (Ronald N. Kostoff, 2002; Ronald N. Kostoff & Martinez, 2005). Sin embargo, se puede adoptar otra perspectiva sobre la normalización por campo. Como es su utilidad en la gestión y evaluación de la investigación para comparar diferentes unidades de investigación (por ejemplo, instituciones, grupos de investigación o investigadores). En el mapeo cuantitativo del perfil multidimensional a nivel macro, meso o micro es necesario calcular este tipo de indicador. Este tipo de normalización no corrige todos los factores de distorsión, pero si rectifica uno de los más importantes. Desde este punto de vista, la normalización por campo tiene un propósito práctico importante.

#### 1.4. Metodologías y modelos para estudiar la ciencia

Esta sección se centra en las diferentes metodologías más utilizadas en bibliometría para analizar y mapear la dinámica de la ciencia. Primero describimos organizaciones que han creado metodologías para reflejar el desempeño, la capacidad tecnológica y el impacto de la actividad económica, científica y de innovación de un conjunto de países en el marco de su contexto socio-económico y científico. Después, analizamos metodologías diseñadas para la visualización de mapas bibliométricos que proporcionan una cartografía de la ciencia.

La OCDE ha desarrollado una metodología para la analizar y evaluar los resultados de la ciencia y la tecnología. México es uno de los países de América Latina que forma parte de los 38 países que integran esta comunidad. Esta metodología constituye una guía para la construcción y uso de indicadores compuestos para los responsables de la formulación de políticas, los académicos, los medios de comunicación y otras partes interesadas (OECD/Eurostat, 2015). La metodología propone un conjunto de indicadores compuestos que permiten comparar y clasificar el desempeño de los países en áreas como la competitividad industrial, el desarrollo sostenible, la globalización, la ciencia y la innovación.

Otra de las metodologías que analizan la ciencia a partir de indicadores de socioeconómicos, bibliométricos y de innovación es la Red Iberoamericana e

Interamericana de Indicadores de Ciencia y Tecnología (RICYT). Creada en abril de 1995, con el apoyo institucional de la Organización de Estados Iberoamericanos, el Instituto de Estadística de la UNESCO, la Secretaría Ejecutiva del Convenio Andrés Bello, el Consejo Caribeño para la Ciencia y la Tecnología y la Comisión para el Desarrollo Científico y Tecnológico de Centro América y Panamá. Su principal objetivo es la implementación de una plataforma que permita monitorear, a través de indicadores científicos, el desempeño de la ciencia y la tecnología en la región Iberoamericana (RICYT, 2019).

Mapear y visualizar la estructura y dinámica de la ciencia ha atraído un gran interés en los últimos años en el campo de los estudios métricos de la información. En este escenario las redes se han vuelto un componente importante para describir y comprender, las conexiones científicas, la colaboración, la identificación de nuevos campos de conocimiento entre otros fenómenos. Para los cientometras el análisis de redes para comprender la ciencia ha sido posible gracias a la serie de herramientas (Pajek, Gephi, VOSViewer, CiteSpace y Science of Science (Sci2)) y metodologías disponibles.

Existen infinitas formas de representar visualmente la literatura científica. Por tanto, es difícil ofrecer una visión completa y sistemática de las metodologías que se han propuesto en la literatura. En términos generales se han reportado metodologías con diferentes propósitos. Décadas atrás la primera iniciativa la tuvo Eugene Garfield con “ISI Atlas of Science”. Proyecto que tuvo como objetivo la visualización y evaluación de campos científicos de manera periódica y sistemática (Garfield, 2009).

Otro ejemplo es el “Atlas of Science” por Katy Borner. Quien propone elementos teóricos en el mapeo científico, los cuales están conceptualizados en su metodología para la exploración y visualización de los resultados de la ciencia y la tecnología (Borner, 2010). Aparecen otros estudios que ofrecen métodos para identificar tópicos emergentes en la ciencia; así como, mapear la estructura y evolución de dominios de conocimiento (Boyack et al., 2005; Boyack et al., 2009).

La tesis de Nees Jan van Eck introduce una nueva metodología para el mapeo científico (Nees Jan van Eck, 2011). La metodología presentada en esta tesis planteó mejores métodos y técnicas para el mapeo científico, especialmente en la normalización para corregir las diferencias entre las unidades de análisis; en la construcción del mapa y la visualización.

Cobo, diseñó una metodología basada en el análisis de co-palabras para la realización del análisis de mapas científicos bajo un marco longitudinal. Esta propuesta puede emplearse para identificar, cuantificar y visualizar la evolución de un dominio científico. Por tanto, la propuesta se centró en redes de co-palabras, analizando aspectos conceptuales de un campo (Cobo Martín, 2011).

Estas propuestas están diseñadas para obtener redes bibliométricas. Las técnicas de mapeo propuestas en estas metodologías permiten identificar por ejemplo según la unidad de análisis seleccionada los más influyentes en la red, las relaciones entre sí o cómo se han desarrollado en el tiempo. Sin embargo, la metodológica propuesta permite el análisis y visualización temporal de datos multidimensionales en la ciencia aplicando inteligencia artificial.

## Capítulo 2: Metodología de la Investigación

### 2.1. Introducción

En este capítulo se exponen los elementos metodológicos que permitieron el análisis de la producción científica de México. La metodología utilizada nos permite obtener varios escenarios gráficos hasta mapear el estado actual y la evolución de los perfiles de desempeño de México, sus instituciones y 22 áreas de investigación. Así como, analizar las asociaciones, estructura y dinámica a diferentes niveles de agregación (global/mundial, regional, países, instituciones, temas), las cuales se muestran en visualizaciones valiosas.

### 2.2. Metodología

En esta sección se aplica la metodología del “Mapeo Neurocomputacional de la Ciencia”. Tomando como estudio de caso el análisis de la producción científica mexicana. Particularmente, esta metodología se explica en detalles en el capítulo 3, porque se puede aplicar en cualquier otro país de la región y del mundo.

#### 2.2.1. Etapa 1. Especificación de la investigación.

El principal objetivo en este estudio es: analizar y visualizar perfiles multidimensionales en la producción científica mexicana. Para ello se seleccionó la colección principal del *Web of Science* (WoS) como fuente de información. Esta fuente está disponible a través del portal de bibliotecas de la Universidad Nacional Autónoma de México (UNAM). Los motivos que nos han llevado a elegir esta base de datos, como fuente de obtención de información bibliográfica para nuestro estudio, obedecen a distintas argumentaciones que se detallan a continuación:

Multidisciplinariedad: Todos los campos de la Ciencia están representados.

Selectividad: Las revistas son escogidas sobre la base de criterios cuantitativos (atendiendo a la cantidad de citas que reciben), y definitivamente seleccionadas según juicio de expertos.

Cobertura total de los contenidos de las más de 10, 000 revistas fuente.

Dirección completa de cada uno de los autores de los trabajos registrados en la base de datos.

Referencias bibliográficas: cada una de las referencias de los artículos se procesan, lo cual resulta una de sus principales prestaciones, pues facilita su utilización para el estudio de la base intelectual y construcción de indicadores.

Se establecieron dos niveles de agregación: a nivel macro considerando a México y a nivel meso sus instituciones y las 22 áreas de investigación.

Además, se consideró incluir las bases de datos del Sistema Nacional de Investigadores (SNI). Estas contienen información sobre el número total de investigadores mexicanos que pertenecen al sistema. El SNI establece criterios de evaluación y otorga un nivel a cada investigador. Además, se clasifican los investigadores en siete áreas de conocimiento; y se calculan indicadores que describen la productividad científica y tecnológica (patentes) (CONACYT, 2018).

### 2.2.2. Etapa 2. Recuperación de datos y/o registros bibliográficos.

La tarea de búsqueda, extracción y descarga de los registros para el estudio se realizó en la interfaz de búsqueda avanzada de la colección principal de la WoS. Se especificó: CU=(Mexico) NOT CU=(New Mexico) y en el campo *Time span* en el *Web of Science*: el período 2000 – 2019.

*Web of Science* solo permite descargar hasta 500 registros. El volumen de la producción científica mexicana entre 2000 y 2019 es igual a 28,5981 registros. La descarga de este gran volumen de registros es muy tardada si se descarga de 500 en 500. Bajo la dirección del Dr. Carrillo, diseñamos una rutina computacional para la descarga automática de los registros del *Web of Science*. Al final obtenemos una base de registros bibliográficos.

### 2.2.3. Etapa 3. Análisis biblio-infor-cienciométrico.

Para diseñar y seleccionar los indicadores del estudio se identificaron en la primera etapa las unidades de análisis o niveles de agregación: México, como país, así como, sus instituciones y 22 áreas de investigación. Después, se definieron las variables y dimensiones que responden a los objetivos del estudio, ver tabla 2:

Tabla 2. Variables y dimensiones para el estudio

<b>Variables</b>	<b>Dimensiones</b>
Producción Científica	Documentos publicados en Web of Science % Documentos publicados en Web of Science Tasa de crecimiento anual del número de documentos producidos Documentos en el 1% de los más citados Documentos en el 10 % de los más citados Documentos publicados en revistas de Q 1
Productividad (eficiencia científica)	Documentos en Web of Science por SNI Documentos en WoS por institución mexicana
Impacto	Impacto real Impacto por Documento Impacto normalizado
Colaboración Científica	Colaboración Internacional
Económico-demográfico	Crecimiento económico Población Capacidades productivas de un sistema económico Prosperidad económica
Innovación	Las patentes por cada 1000 habitantes
Liderazgo	Autores para correspondencia en un artículo
Esfuerzo científico	Ventajas comparativas en un dominio científico

La siguiente tarea consistió en seleccionar el conjunto de indicadores que miden las variables previamente identificadas (ver tabla 2). Un total de 18 indicadores seleccionados caracterizaron el perfil cienciométrico de la producción científica mexicana. Seis indicadores de producción; tres indicadores bibliométricos de

impacto real; un indicador de colaboración científica, dos indicadores de productividad/eficiencia, un indicador de innovación, tres indicadores socioeconómicos, uno que refleja esfuerzo científico y otro de liderazgo. A continuación se conceptualizan:

### 2.2.3.1. Indicadores de producción y productividad

Los indicadores de producción tienen una función descriptiva, en la cual estos están encaminados a producir información sobre el estado del volumen del sistema científico objeto de estudio. La productividad se interpreta como eficiencia y el rendimiento del sistema como efectividad, términos que provienen de la economía. La interpretación económica de la ciencia se basa en la productividad de la investigación (eficiencia) y en el rendimiento (efectividad) a través del cálculo de indicadores que tomen en cuenta las medidas de inversión en la investigación (Abramo, Andrea, Angelo, & Soldatenkova, 2016). Específicamente, en la investigación se determinó incluir los siguientes indicadores de producción y productividad:

Ndoc: Total de artículos publicados por años, áreas de investigación, países, revistas.

%Ndoc: Porcentaje del total de artículos publicados por años, áreas de investigación, países; los cuales miden el volumen de la producción científica mexicana durante el período de estudio.

Para medir la productividad/eficiencia de la producción mexicana, el presente estudio consideró dos indicadores independientes de tamaño, con el objetivo de realizar comparaciones más justas: Índice de productividad nacional y el Índice de productividad institucional.

NSP: Índice de productividad nacional (*National Scientific Productivity*). Ndoc Wos/SNI. Es el cociente del número de documentos indexados en WoS por autores mexicanos, entre el número de investigadores en el Sistema Nacional de Investigadores (SNI).

Es importante señalar que, el indicador es una estimación de la productividad nacional, porque hay artículos científicos producidos en México por investigadores que no pertenecen al Sistema Nacional de Investigadores. Sin embargo, el número de artículos cuyo autor no es miembro de SNI, constituyen un pequeño porcentaje de la producción total del país.

ISP: Índice de productividad institucional (*Institucional Scientific Productivity*). Ndoc por instituciones mexicanas/SNI.

TCWoS: Tasa de crecimiento anual del número de documentos producidos. Valor promedio de las tasas de crecimiento anuales del número de documentos producidos en los últimos años comprendidos en el periodo: 2000-2019

% Top 1%: Porcentaje de Excelencia: porcentaje de documentos en el 1% de los más citados por cada campo de investigación, año y tipo de documento.

% Top 10%: Alto desempeño (*High performance*): porcentaje de documentos en el 10% de los más citados por cada campo de investigación, año y tipo de documento.

%Q1: porcentaje de documentos con al menos un autor mexicano publicados en una revista que se encuentra en el cuartil 1.

### 2.2.3.2. Indicadores de impacto y visibilidad real

Los indicadores basados en las citas ofrecen varias interpretaciones: como una medida del impacto científico de la investigación; de prestigio académico para otorgar premios y honores; también reflejan la calidad, relevancia e impacto de un documento en la obra de otro, para de esta forma demostrar que, las citas se realizan a los aportes científicos(Waltman, 2016).

Se consideraron algunos de los indicadores basados en las citas que ofrece el *Web of Science* para estudiar, analizar y comparar la producción científica mexicana con otros países. Los indicadores que permitieron analizar el impacto de la ciencia mexicana son:

Ncit: Total de citas recibidas

Ncitas/Doc: Relación de citas por documento

CNCI: Category Normalized Citation Impact: se calcula dividiendo el recuento real de citas recibidas por la tasa de citas esperadas para documento con el mismo tipo, año de publicación y área temática. Normaliza el impacto, tomando en cuenta la categoría temática, el tipo de documento y la fecha de publicación.

Cuando el CNCI tiene un valor igual a uno, representa el rendimiento a la par con el promedio mundial, los valores superiores a uno se consideran superiores al promedio y los valores inferiores a uno se consideran inferiores al promedio. Un valor CNCI de dos se considera el doble del promedio mundial. Este indicador puede calcularse de varias formas, las cuales se mencionan a continuación(Clarivate Analytics, 2018):

- Para un solo documento que es asignado a un área temática:

$$CNCI = \frac{c}{e_{ftd}}$$

- A un artículo se le asignan múltiples categorías temáticas. El CNCI se puede representar como el promedio de las proporciones de citas reales a esperadas para cada área temática:

$$CNCI = \frac{\sum \frac{c}{e_{f(n)td}}}{n} = \frac{\frac{c}{e_{f(1)td}} + \frac{c}{e_{f(2)td}} \dots + \frac{c}{e_{f(n)td}}}{n}$$

- Para un grupo de artículos, el valor CNCI es el promedio de los valores para cada uno de los papeles, representado como:

$$CNCI_i = \frac{\sum_i CNCI_{each\ paper}}{p_i}$$

Donde: e = tasa de citas esperadas, c = total de veces citado, p = número de artículos, f = campo o área temática, t = año, d = tipo de documento, n = número categorías a las que se asigna un documento y i = la unidad de análisis que se está evaluando (institución, país / región, persona, etc.).

CNCI es un indicador ideal para la evaluación comparativa en todos los niveles organizativos (autor, institución, región, etc.). Este indicador permite comparar unidades de análisis con diferentes tamaños y procedentes de disímiles campos de investigación (Waltman & van Eck, 2015). En la identificación de subconjuntos de documentos con un impacto promedio superior a la media mundial. Para determinar el desempeño de investigadores noveles en comparación con los establecidos. Puede ser un indicador para evaluar aspirantes a una plaza académica. El CNCI es útil a las agencias de financiamiento como un indicador cuantitativo para monitorear el desempeño de proyectos financiados o evaluar un grupo de investigación que solicita financiamiento.

#### 2.2.3.3. Colaboración científica

Entre los elementos de mayor importancia en los estudios de I+D, es la colaboración científica. Las redes que se forman en el interior de los niveles, constituyen el espacio donde se generan, intercambian y emplean los conocimientos que conducen a la innovación. Los estudios bibliométricos sobre la colaboración científica son complejos debido a los aspectos metodológicos que se deben considerar: la normalización y los tipos de conteos.

En este estudio se utilizó el conteo completo como método de recuento. Se asignó cada documento a todas y cada una de las instituciones o países firmantes del mismo.

%CI: % Colaboración Internacional: Porcentaje de publicaciones que tienen coautores internacionales.

#### 2.2.3.4. Indicadores económicos-demográficos

Los indicadores se descargaron de las bases de datos del Banco Mundial<sup>1</sup> y del *Atlas of Economic Complexity*<sup>2</sup>. Básicamente, son medidas estadísticas referentes a la dimensión demográfica y económica de una sociedad humana.

Pob 2019: La población total se basa en la definición de facto de población, que cuenta a todos los residentes independientemente de su estado legal o ciudadanía anualmente (World Bank Group, 2021).

PIB 2019 per cápita: es el producto interno bruto dividido por la población. El PIB es la suma del valor agregado bruto de todos los productores residentes en la economía más todo impuesto a los productos, menos todo subsidio no incluido en el valor de los productos. Se calcula sin hacer deducciones por depreciación de

---

<sup>1</sup> Datos de libre acceso del Banco Mundial. <https://datos.bancomundial.org/>

<sup>2</sup> The Atlas of Economic Complexity. <https://atlas.cid.harvard.edu/>

bienes manufacturados o por agotamiento y degradación de recursos naturales. Datos en US\$ a precios actuales. (World Bank Group, 2021).

EI (*Complexity Index* o Índice de Complejidad Económica): es una medida holística de las capacidades productivas de los grandes sistemas económicos, generalmente ciudades, regiones o países. En particular, la EI busca explicar el conocimiento acumulado en una población y que se expresa en las actividades económicas presentes en una ciudad, país o región.

El EI fue desarrollado por Cesar A. Hidalgo, del MIT Media Lab y Ricardo Hausmann, de la Kennedy School of Government de la Universidad de Harvard. Los datos de EI están disponibles en *The Observatory of Economic Complexity* y en el sitio del *Atlas of Economic Complexity* (Hidalgo & Hausmann, 2011).

GCI (*Global Competitiveness Index* o Índice de Competitividad Global): mide en un país determinado un conjunto de instituciones, políticas y factores que establecen los niveles sostenibles de prosperidad económica actual y de mediano plazo. Este índice es publicado anualmente en el Informe de Competitividad Global por el Foro Económico Mundial (Schwab, 2019).

#### 2.2.3.5. Indicadores de innovación, liderazgo y actividad científica

C INV (Coeficiente de Inventiva): El coeficiente de invención se refiere a las patentes solicitadas por los residentes cada 1000 habitantes.

CorrA(2008-2020): Autor para correspondencia, porcentaje de publicaciones en las que el autor elegido se encuentra en la posición de recibir las notificaciones sobre los documentos publicados a partir de 2008.

RCA (*Revealed Comparative Advantage*): permite evaluar las ventajas comparativas en un dominio científico para un país con respecto al mundo. Este indicador fue propuesto por Balassa en 1977 (Balassa, 1977). Después se adoptó en el campo de la cienciometría (Chuang, Lee, Hung, & Lin, 2010).

Después de seleccionar los indicadores, utilizamos Bibliometrix R-package (Aria & Cuccurullo, 2017) de Rstudio (R Core Team, 2016) para calcular los siguientes indicadores bibliométricos a nivel macro: total de artículos publicados por años; total de citas recibidas; relación de citas por documento.

Luego, en *InCitesB&A* se obtuvieron otros indicadores; los cuales no pueden ser calculados con *Bibliometrix*. Por ejemplo: *Category Normalized Citation Impact*, porcentaje de la producción científica en el 1% y 10% de lo más citados, por ciento de la producción en revistas de primer cuartil a nivel macro y meso.

Se establecieron filtros dentro de la herramienta atendiendo a las siguientes variables: organización, país, WoS como esquema de clasificación, el período 2000-2019 y especificar indicadores de producción, impacto, colaboración, excelencia y alto desempeño por instituciones.

En el caso de las áreas de investigación mexicanas, se filtró primeramente el esquema del *ESI*, período, el país y los indicadores. Después, se descargó un documento con extensión CSV con los indicadores.

Los indicadores de productividad (NSP e ISP) se calcularon en *notebooks* de *GoogleColab* utilizando Python como lenguaje de programación.

Al final de la etapa 3, se obtuvo una base de datos de indicadores cientimétricos calculados con *BibliometrixR*, *Python* y se incluyó el conjunto de indicadores obtenidos con *InCitesB&A* de WoS.

#### 2.2.4. Etapa 4. Análisis neurocomputacional y visualización.

Primero, se crearon *notebooks* en *GoogleColab* para cada nivel de agregación o unidad de análisis (figura 9). Luego, se programaron los códigos en *Plotly* (*Python*), obteniéndose diferentes escenarios gráficos. Los cuales mostraron estudios unidimensionales y bidimensionales la evolución de los tipos de producción científica mexicana, la comparación de México y países latinoamericanos considerando indicadores de producción e impacto.

```

from google.colab import drive
drive.mount('/content/gdrive')

Resultado de /content/gdrive

Evolución de la PC MX en WoS y SCOPUS 2000-2019

import pandas as pd
PCWOSySCOPUS = pd.read_csv('/content/gdrive/My Drive/Doctorado Copia/tesis PhD/200-2019/InCites Mex/WoS y SCOPUS.csv')
PCWOSySCOPUS.head()

Publication Year  Web of Science Documents  Scopus
0                2000                    6469   6545
1                2001                    7134   7013
2                2002                    7370   7791
3                2003                    8006   8018
4                2004                    8006   8056

! pip install plotly==1.11

Requirement already satisfied: plotly==1.11 in /usr/local/lib/python3.4/dist-packages (4.1.1)
Requirement already satisfied: numpy==1.12.1 in /usr/local/lib/python3.4/dist-packages (from plotly==1.11) (1.12.1)
Requirement already satisfied: six in /usr/local/lib/python3.4/dist-packages (from plotly==1.11) (1.11.0)

! PCWOSySCOPUS['Web of Science Documents'].values

array([ 6469,  7134,  7370,  8006,  8235,  8598, 10432, 12172, 13095,
        13284, 14019, 14429, 14671, 17076, 17843, 19354, 20731, 22042,
        22720, 23679])

```

Figura 9. *Notebook* de la producción científica mexicana en *GoogleColab*, programado en *pyhton* y utilizando *plotly*.

ViBlioSOM 2.0 sirvió para obtener una caracterización multidimensional y temporal del perfil cientimétrico de México, sus instituciones y áreas de investigación. La red neuronal permitió obtener mapas multidimensionales, neuro longitudinales y de órbitas. Estos mapas están integrados por mapas de clustering y mapas de componentes.

Particularmente, los mapas multidimensionales permitieron comparar a México con países latinoamericanos; identificar el conjunto de países con similar desarrollo socioeconómico y luego contrastar sus perfiles cientimétricos. Así como, comparar el perfil de las instituciones mexicanas más productivas.

Los mapas neuro longitudinales facilitaron el análisis de la evolución del perfil cientimétrico de las áreas de investigación para los años 2017, 2018 y 2019. Con el mapa de órbitas se logró visualizar la dinámica del perfil cientimétrico de México en el período 2000-2019.

## **Capítulo 3. Análisis y discusión de los resultados**

### **3.1. Introducción**

Este capítulo se presenta la metodología diseñada en el marco de esta investigación para analizar y visualizar la evolución de perfiles cientométricos multidimensionales. A la cual hemos llamado: Mapeo Neurocomputacional de la Ciencia.

En esta metodología se ha integrado técnicas de análisis bibliométrico y visualización de información, basadas en aprendizaje de máquina (*machine learning*) e inteligencia artificial. Tiene el objetivo de analizar y visualizar la configuración y dinámica de datos multidimensionales, y aportar visualizaciones que facilitan el descubrimiento de conocimientos.

“Mapeo Neurocomputacional de la Ciencia” está organizada en cuatro etapas, las cuales ofrecen diferentes análisis. La metodología permite realizar, desde análisis unidimensionales, con diagramas de dispersión 2-D que agregan forma, volumen o color, hasta análisis temporal de múltiples indicadores cientométricos. Este conjunto de indicadores caracterizan el perfil cientométrico de una unidad de análisis.

### **3.2. Mapeo Neurocomputacional de la Ciencia**

En esta sección presentamos un modelo de relaciones de los elementos metodológicos que nos servirá para describir nuestra metodología en la sección 3.2. Describiendo el orden lógico las etapas y tareas, que son las siguientes (figura 6):

1. Especificación de la investigación
2. Recuperación de datos y/o registros bibliográficos
3. Análisis bibliométrico
4. Análisis neurocomputacional y visualización

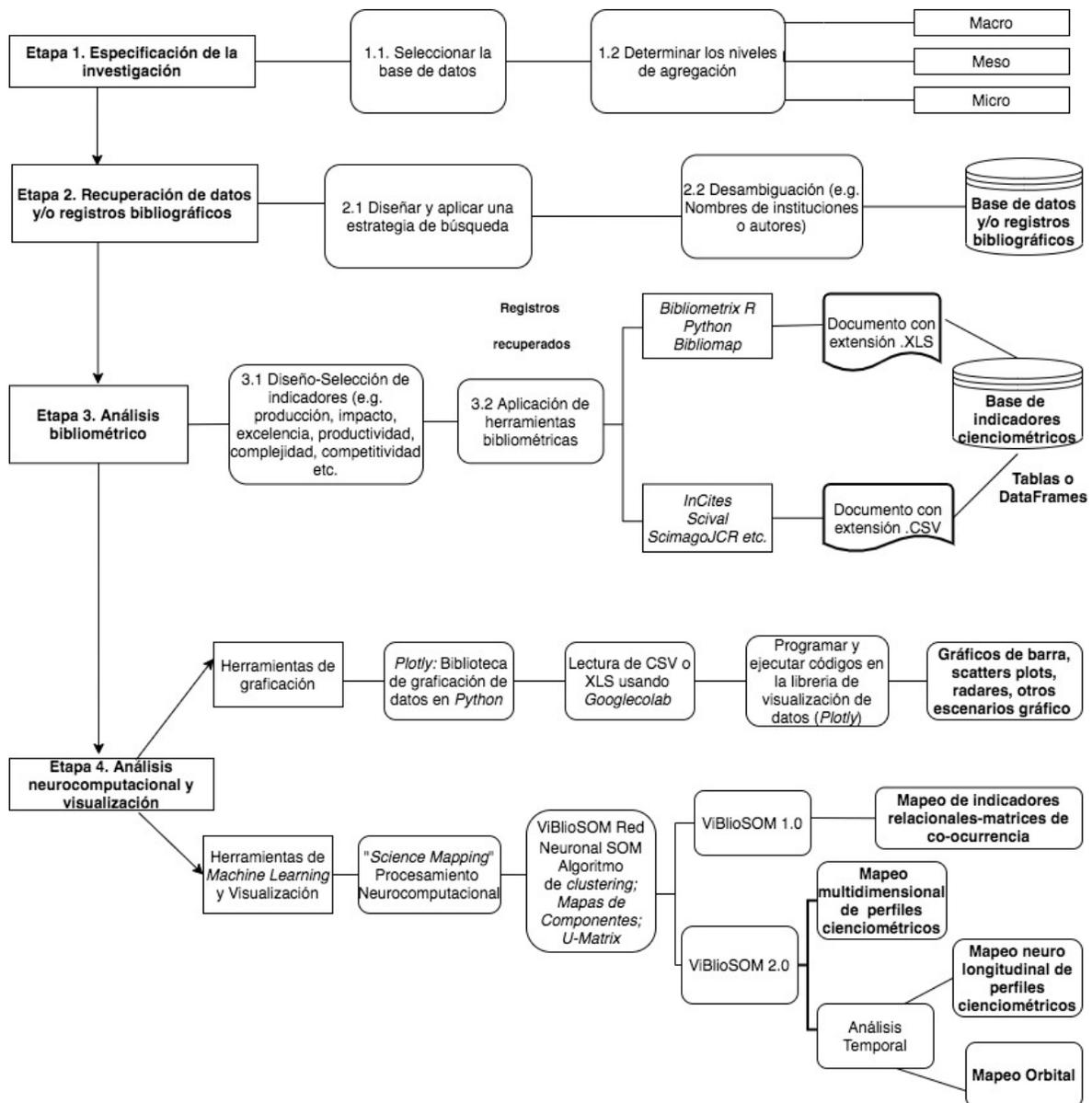


Figura 6. Modelo de la metodología.

### 3.2.1. Etapa 1: Especificación de la investigación

En la especificación de la investigación, se debe definir el objetivo de la investigación o análisis y su alcance. Después, la selección de la base de datos y los niveles de agregación están determinados por los objetivos y el alcance previamente definidos. Particularmente, la selección de la base de datos debe estar basada en el cobertura temática y grado de actualización en los datos para ofrecer robustez en los datos.

Los niveles de agregación o unidades de análisis se jerarquizan de la siguiente forma: macro (nivel mundial/global, regional, país), meso (instituciones,

clasificaciones temáticas generales), micro (autores, grupos, departamentos, sub-categorías disciplinarias). Para determinar el nivel de agregación al que se trabajará es importante conocer el alcance de la investigación que se realizará, el volumen de datos disponibles y el grado de ambigüación de la información.

También, es necesario incluir otras fuentes de información cuando el objetivo del estudio es construir y visualizar perfiles multidimensionales. Los cuales van a reflejar la complejidad de la producción científica. Para ello, se pueden utilizar las bases de datos del Sistema Nacional de Investigadores (SNI), las bases de datos del Banco Mundial, el *Atlas of Economic Complexity* por solo mencionar algunas.

### 3.2.2. Etapa 2. Recuperación de datos y/o registros bibliográficos

El diseño y aplicación de la estrategia de búsqueda a emplear incluye considerar las condiciones que la base de datos seleccionada en la etapa 1 (ver figura 6) establece para la recuperación de los registros. En este caso, usar la búsqueda avanzada de la base de datos ofrece un grado mayor de especificidad.

En esta interfaz se pueden combinar los puntos de acceso (campos de la base de datos) con los operadores booleanos y paralelamente se puede especificar el período en el que se desea analizar la producción científica. Los registros bibliográficos se descargan con la salida en esta etapa.

Con los registros bibliográficos descargados se procede a la desambiguación del nombre de instituciones o nombre de autores. Esta tarea debe eliminar el 95% del nivel de ambigüedad cuando se atribuyen registros bibliográficos específicos al nombre de las instituciones o nombre de autores. La salida final en esta etapa es la base de registros bibliográficos.

### 3.2.3. Etapa 3. Análisis biblio-infor-cienciométrico

El diseño y selección de los indicadores de la investigación considera los objetivos de la investigación y el nivel de agregación del estudio (unidades de análisis). Se definen en la etapa 1, e.g., regiones, países, instituciones, áreas de investigación, revistas etc. Teniendo identificado el nivel de agregación, se determinan las variables que resultan de interés para la investigación.

En el marco de esta metodología, una variable se define como una característica, cualidad o medida que puede sufrir cambios. Se trata de identificar y señalar con precisión los componentes, dimensiones o factores que teóricamente integran la variable. De igual manera, se establecen los indicadores de cada dimensión (Hernández-Sampieri, Fernández, & Baptista, 2014). En esta etapa identificamos variables medibles de la producción científica, dimensiones y sus indicadores para describir y explicar sus cambios.

La batería de indicadores deben capturar diferentes dimensiones de la producción científica los cuales caracterizan el perfil cienciométrico de la unidad de análisis escogida en la primera etapa, por ejemplo: producción, impacto real, colaboración científica, productividad etc. Además, en este conjunto se pueden incluir indicadores socioeconómicos, indicadores de innovación etc con el objetivo de relacionar los recursos/capacidades con las salidas de la ciencia.

Una vez seleccionado y definido la batería de indicadores se aplican herramientas bibliométricas para calcular u obtener estos. La entrada, en esta etapa, es la base de registros bibliográficos obtenidos en la etapa anterior.

En este caso el experto tiene dos opciones: a) Diseñar las líneas de código en *R*, *Python*, *Bibliomap* u otras herramientas para procesar los registros bibliográficos y calcular los indicadores bibliométricos; y b) obtener automáticamente los indicadores de las herramientas bibliométricas en las bases de datos: *InCitesB&A* de *WoS*, *SciVal* de *Scopus*, *ScimagoJCR*.

La tercera opción combina ambas posibilidades cuando los indicadores que proveen las herramientas bibliométricas no son suficientes dado los objetivos de la investigación. De esta forma, se calculan un conjunto mayor de indicadores.

La obtención de los indicadores en las herramientas bibliométricas como *InCitesB&A* o *SciVal* es automáticamente. El experto especifica determinados filtros en la interfaz de la herramienta. Los filtros tienen que estar de acuerdo al nivel de agregación que se va a analizar la producción científica. Este tipo de herramienta es excelente cuando queremos realizar estudios a nivel meso y micro (Vera-Baceta, Thelwall, & Kousha, 2019).

Diferentes herramientas bibliométricas ofrecen la posibilidad de calcular las principales medidas bibliométricas. Proponemos utilizar *Bibliometrix R-package* (Aria & Cuccurullo, 2017) de *Rstudio* (R Core Team, 2016) para calcular los indicadores (producción científica anual, total de citas por año, autores más citados, índice H, ley de Lotka). La figura 7 muestra el esquema de pasos para trabajar con el software.

Este paquete permite calcular las principales medidas bibliométricas y generar tablas con los indicadores (total de artículos, período, tasa de crecimiento del número de artículos publicados, promedio de citas por artículo, total de autores, autores por artículo, índice de colaboración etc). Además, la función de citas genera la tabla de frecuencias con las referencias más citadas o los autores más citados (de referencias), y también calcula el índice h (Hirsch, 2005) para los autores.

La función *normalizeSimilarity* facilita al usuario normalizar los datos antes de crear una red de acoplamiento bibliográfico, co-citación y co-ocurrencia. En la investigación cuantitativa, cuando se aplica una medida de similitud directa a los datos de co-ocurrencia, el objetivo es corregir los datos por diferencias en el número de ocurrencias de objetos (Nees Jan Van Eck & Waltman, 2009).

*Bibliometrix R-package* tiene implementadas varias medidas de normalización (la fuerza de asociación; índice de inclusión; coeficiente de Jaccard y coeficiente de similitud de Salton), sugerimos usar la fuerza de asociación para normalizar los datos en redes de co-ocurrencia (Nees Jan Van Eck & Waltman, 2009).

*cocMatrix*, es una función para obtener las redes bibliométricas (acoplamiento bibliográfico, co-ocurrencia, co-citación y colaboración). Por su parte, *biblioNetwork* es la función general para crear una matriz rectangular (Documento × Atributo). Un atributo es un elemento de información asociado al documento y almacenado en una etiqueta de campo dentro de la tabla de datos (por ejemplo, autores, fuente de publicación, palabras clave, referencias citadas, afiliaciones). Las redes que se obtienen pueden luego ser importadas otros sistemas de software como VOSviewer.

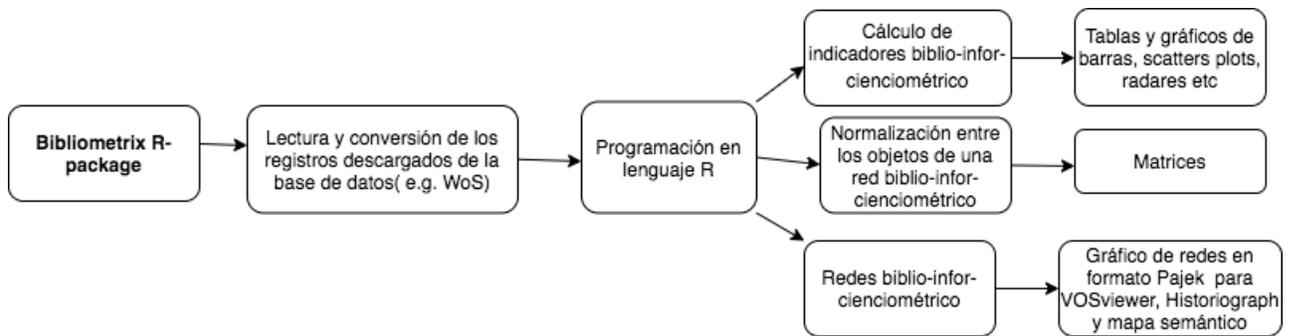


Figura 7. Análisis biblio-infor-cienciométrico de la producción científica con Bibliometrix R-package.

Al final, se obtiene una base de datos de indicadores científicos. Esta base contiene los .CSV y .XLS o en inglés llamados *DataFrame*. Los cuales serán la fuente de entrada en la siguiente etapa.

#### 3.2.4. Etapa 4: Análisis neurocomputacional y visualización

En el análisis y visualización se aplican herramientas de visualización y de *machine learning*. El objetivo de esta etapa es explorar, comprender, analizar y visualizar la información valiosa y las relaciones que pueden estar ocultas dentro de los datos. La base de indicadores científicos, es la entrada en esta etapa.

Las herramientas de visualización de datos brindan conocimientos que los datos por sí solos no pueden. *Python* es excelente para la exploración y el análisis de datos, y todo gracias a la compatibilidad con bibliotecas como *numpy*, *pandas*, *matplotlib* y *plotly*. Esta última, es una de las bibliotecas más utilizadas en *Python*.

Con *Plotly* es posible crear gráficos como: diagramas de líneas, diagramas de dispersión, radares, diagramas de barras, diagramas de caja, histogramas, mapas de calor o *heatmaps*, etc. Estas visualizaciones describen perfectamente los análisis unidimensionales o biparamétricos. Por tanto, es la librería que usaremos en esta etapa.

El siguiente paso es la creación de *notebooks* en *GoogleColab* para obtener gráficos y visualizaciones. Esta tarea implica conocimientos de programación en *Python* para programar la lectura de los .CSV y .XLS(*DataFrame*) utilizando *GoogleColab*. Luego, se asigna un orden a las investigaciones contenidas en los *notebooks*. Este orden consiste en calcular los indicadores según el grado de complejidad considerando los objetivos planteados en la primera etapa. Al final, los gráficos y visualizaciones tienen un orden lógico.

Después de obtener los gráficos identificamos que, algunas de las cuestiones reveladas por el análisis unidimensional y biparamétrico se comprenden mejor desde una perspectiva multidimensional. Por ello, la siguiente tarea es aplicar herramientas de *machine learning*.

El procesamiento de los indicadores científicos es con tecnología basada en redes neuronales artificiales, específicamente, técnicas de mapeo autoorganizado. La aplicación de este tipo de tecnología permite la caracterización

multidimensional y temporal del perfil cientimétrico de la unidad de análisis que se está investigando.

El procesamiento neurocomputacional se realiza con el sistema de software ViBlioSOM 2.0 basado en el algoritmo SOM. El algoritmo SOM toma datos bibliométricos multidimensionales y los proyecta, de forma no lineal, a un mapa bidimensional, preservando la estructura topológica (los datos cercanos en el mapa significa proximidad en el espacio multidimensional original). Esta tarea se presenta en el diagrama de la figura 8.

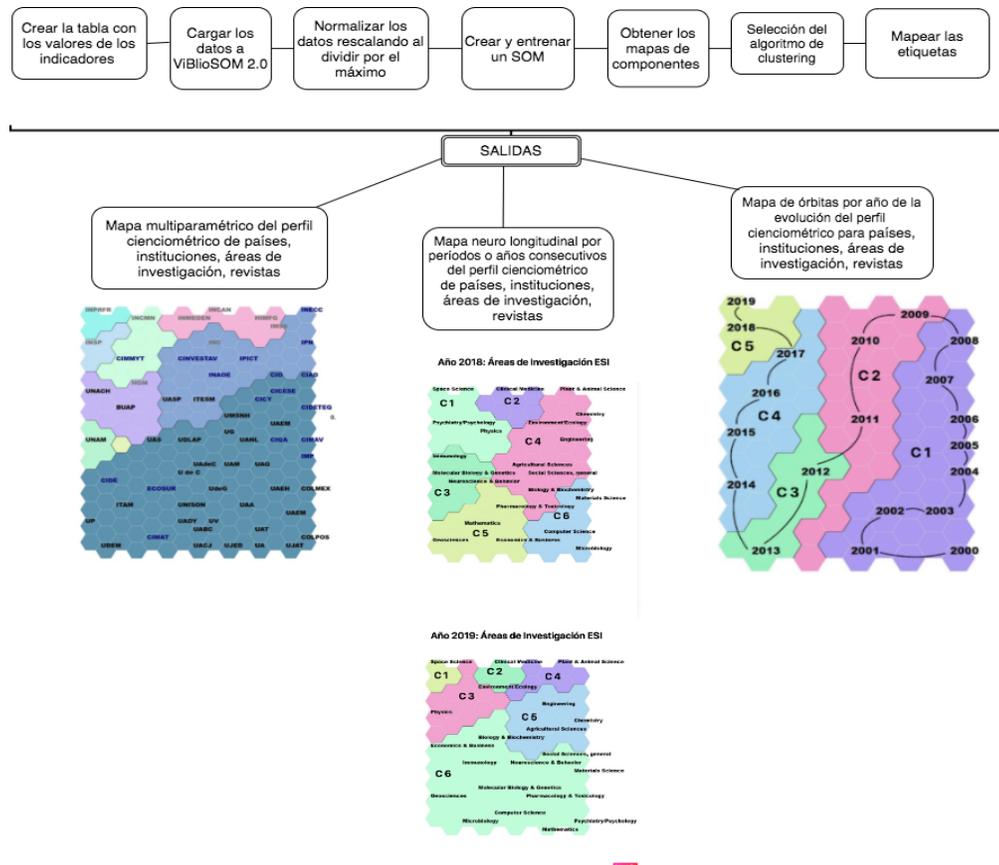


Figura 8. Pasos para el procesamiento neurocomputacional de datos multidimensionales que caracterizan el perfil cientimétrico de un país, institución o campo de conocimiento.

El primer paso en esta tarea es crear la matriz con los datos de los indicadores. Luego, importar esta matrix a ViBlioSOM 2.0, estos datos deben ser normalizados, se rescalan al dividir por el máximo. Después, se entrenan las neuronas hasta obtener los mapas U-Matrix, componentes y el mapa de *clustering*. El mapa U-Matrix y de componentes deben ser explorados primero. Luego, creamos el mapa de clustering.

### Mapa U-Matrix

En la pestaña "Mapas" de ViBlioSOM 2.0 se muestra el mapa U-Matrix. Este mapa es muy útil para visualizar las fronteras de los clusters detectados en el proceso de entrenamiento. En este caso, el rojo implica una mayor distancia en el

espacio de entrada entre las unidades vecinas en el mapa y representa las fronteras entre los cúmulos de datos. Después, los mapas de componentes explican la separación entre los grupos de datos.

### **Mapas de componentes**

Para cada componente de los vectores de referencia se crea un mapa. Los componentes de los vectores de peso corresponden a indicadores, cada mapa de componentes corresponde a un indicador. Para generar un mapa de componentes específico, el color de la neurona se asigna de acuerdo con el valor de ese componente en el peso asociado a la neurona. En el modelo RGB, el color pasa del verde al rojo, pasando por el amarillo. Normalmente, el verde se asigna a los valores más bajos y el rojo a los más altos, pero esto se puede invertir si es conveniente.

### **Mapa de conglomerados o *clustering***

Un conglomerado o *clustering*, es la partición de un conjunto en un conjunto de grupos, de modo que los elementos de un conglomerado son similares entre ellos y diferentes de los elementos del resto. En esta etapa, cada neurona pertenece exactamente a un grupo. En el mapa resultante cada cluster tiene un color y las neuronas se colorean de acuerdo al grupo al que pertenecen. Esos datos dentro de un grupo comparten características generales que pueden conceptualizarse para sacar conclusiones sobre los datos (figura 8).

ViBlioSOM 2.0 tiene disponible varios criterios de clustering. Las opciones disponibles son: la distancia promedio entre los datos de un cluster con los datos del otro (*average-linkage*), la distancia entre los datos más cercanos de los dos grupos (*single-linkage*), la distancia entre los datos más distantes de los grupos (*complete-linkage*) y la cuarta opción define la distancia entre dos grupos como el incremento en la suma de las distancias al cuadrado entre los datos al unir los dos clusters (Ward). En este trabajo evaluamos las cuatro opciones y optamos por la versión *Ward* (Villaseñor et al., 2017).

Para identificar el número óptimo de *clusters* en el mapa de *clustering*, el software, tiene implementado tres indicadores: *Silhouette Coefficient*, *Calinski-Harabasz Index* y *Davies-Bouldin Index*. Basados en los datos que estamos analizando optamos por seleccionar el número óptimo de clusters utilizando el índice de *Davies-Bouldin* (Davies & Bouldin, 1979).

Concluido el proceso de entrenamiento SOM se generan diferentes salidas. A partir de ese momento se puede usar la red y los pesos almacenados dentro de cada neurona para realizar diferentes análisis: análisis multiparamétricos de los perfiles y análisis temporal multidimensional. Las salidas que se pueden obtener son:

#### *a) Mapeo de indicadores relacionales-matrices de co-ocurrencia*

La entrada al sistema de software ViBlioSOM 2.0 son matrices de co-ocurrencia (términos-términos, términos-años, autores-años, autores-autores, etc.). En este nivel se obtienen mapas que permiten estudiar la estructura de las áreas de investigación en la ciencia, empleando las técnicas de análisis de co-ocurrencia; así como, influencia o cercanía entre las variables de análisis (Guzmán Sánchez, 2009).

### *b) Mapeo multidimensional de perfiles cientiométricos*

El mapeo multidimensional de los perfiles cientiométricos se obtienen con el sistema de software ViBlioSOM 2.0. El software realiza una proyección no lineal desde un espacio multidimensional a un mapa bidimensional (mapa de componentes y clustering) conservando la estructura topológica. Esto significa que dos puntos cercanos en el espacio multidimensional se asignan a lugares cercanos en el mapa. Esto está garantizado por la proyección preservadora de la topología realizada por el algoritmo SOM (Kohonen, 1990).

Particularmente, el mapa de clustering es el resultado del proceso de entrenamiento iterativo de la red neuronal, mediante el cual la red aprende a proyectar patrones similares en ubicaciones cercanas (hexágonos) en el mapa. La proyección proporciona una representación de la distribución del conjunto de datos en el espacio multidimensional en una cartografía 2D que permite la visualización de los datos. Se pueden realizar comparaciones entre unidades científicas con un desempeño científico similar (Villaseñor et al., 2017).

### *c) Mapeo neuro longitudinal de perfiles cientiométricos*

Mediante el uso de una serie de mapas ordenados cronológicamente, se pueden observar los cambios de un año a otro, identificar tendencias y este tipo de visualización puede convertirse en una herramienta de predicción. La técnica utilizada en el mapeo neuro longitudinal combina, el aprendizaje automático con la técnica de mapeo longitudinal de Garfield. Esta combinación es útil para el análisis temporal (Garfield, 2004).

En el mapeo neuro longitudinal aplicamos la misma serie de pasos que se describen en el análisis multiparamétrico (Villaseñor et al., 2017). Durante el proceso de entrenamiento de la red neuronal SOM para obtener los mapas neuro longitudinales se refinan los datos para cada período o año (fase de refinamiento) en el SOM. Esta fase sigue los siguientes pasos:

Cargar los pesos del periodo anterior como condición inicial

Asignar un valor pequeño ( 8 o 10) a sigma en la función de vecindad, función que se activa en el botón de parámetros de configuración del ViBlioSOM 2.0

Ejecutar la décima parte de las iteraciones del periodo de referencia. Para el caso del SOM Básico, asignar un valor inicial pequeño para el factor de aprendizaje.

### *d) Mapeo orbital*

El mapeo de trayectorias permite visualizar en un solo mapa la dinámica del perfil de desempeño cientiométrico. A diferencia del mapeo neuro longitudinal que genera un mapa de clustering y componentes para cada año o período de análisis.

En este caso, los perfiles de desempeño de las unidades científicas se modelan con un vector de n dimensiones. Cada dimensión corresponde a un indicador bibliométrico (por ejemplo: impacto normalizado, producción de excelencia,

colaboraciones, entre otros). Al considerar mediciones anuales de los indicadores, obtenemos una secuencia de vectores en un espacio multidimensional.

El procedimiento para obtener un mapa de trayectorias con la evolución del perfil multidimensional para una unidad científica son una extensión de los pasos en la figura 8. Adicionalmente, es necesario antes de crear y entrenar el SOM suavizar los datos con una media móvil de 3 o 4 años. Al final se obtiene un mapa de clustering, el cual visualiza la trayectoria del perfil multidimensional cienciométrico de un año a otro.

### **3.3. Comprobación de la Metodología: Mapeo Neurocomputacional de la Ciencia. Su aplicación a la ciencia mexicana.**

En esta sección se analizan y disuten los resultados que se presentan sobre la producción científica en el WoS durante 2000-2019. Se visualizaron a través de gráficos de barras para analizar la evolución de la producción científica mexicana en comparación con la producción en Scopus, la dinámica de los miembros del Sistema Nacional de Investigadores. Las gráficas de líneas y *scatters plot* permitieron comparar la producción per capita, excelencia y liderazgo de México en el contexto latinoamericano y mundial.

Se identifican y grafican las instituciones mexicanas más productivas, los países que más colaboran con Mexico. Además, se realizan comparaciones bidimensionales y tridimensionales de la productividad, impacto normalizado, excelencia y visibilidad entre las instituciones. Después, los gráficos de radar facilitaron el análisis de las áreas de investigación mexicana con mayor contribución en la producción nacional y mundial.

Las herramientas de *machine learning* y visualización nos permitieron representar datos cienciométricos multidimensionales. La primera aplicación se presenta en la subsección 3.4.2, para analizar, comparar y visualizar los perfiles cienciométricos multidimensionales de los diez países latinoamericanos más productivos en WoS. Luego, se aplicó la misma técnica para identificar los países que tienen un perfil económico-demográfico similar a México y después comparar sus perfiles cienciométricos multidimensionales.

La combinación de la técnica de aprendizaje automático con la técnica de mapeo longitudinal de Garfield clásica, resulta aquí útil para el análisis temporal de los perfiles cienciométricos multidimensionales. Particularmente, en la sección 3.6.3 presentamos una secuencia de mapas longitudinales neuronales que, visualizan la evolución de 2017 a 2019, de los perfiles cienciométricos multidimensionales de las 22 áreas de investigación mexicanas.

Por último, en la metodología diseñada introducimos una nueva técnica de mapeo, basada en redes neuronales artificiales, a la cual hemos llamamos mapeo orbital. En este caso, la red neuronal permitió mapear la evolución del perfil cienciométrico multidimensional de México, durante 2000-2019; considerando siete indicadores cienciométricos.

### 3.4. Mapeo de la evolución de la producción científica mexicana durante el período 2000-2019.

México ha triplicado la producción anual durante los últimos 20 años en el WoS (Figura 10). Particularmente, la producción anual de México en el *Web of Science* durante el período 2000-2019 pasó, de 6469 documentos producidos en el 2000, hasta 24,394 documentos en 2019, a una tasa promedio anual de crecimiento del 6%, alcanzando un volumen acumulado de 285,981 documentos.

En Scopus, el volumen de documentos también se incrementó pasando de 6546 en el 2000 a 27,955 documentos en el 2019 con una tasa promedio anual de crecimiento del 7%, acumulando un total de 324,921. En el período la producción científica de México ha aumentado en WoS y Scopus a una tasa promedio anual del 6% y 7%.

Evolución de la producción mexicana en WoS y Scopus

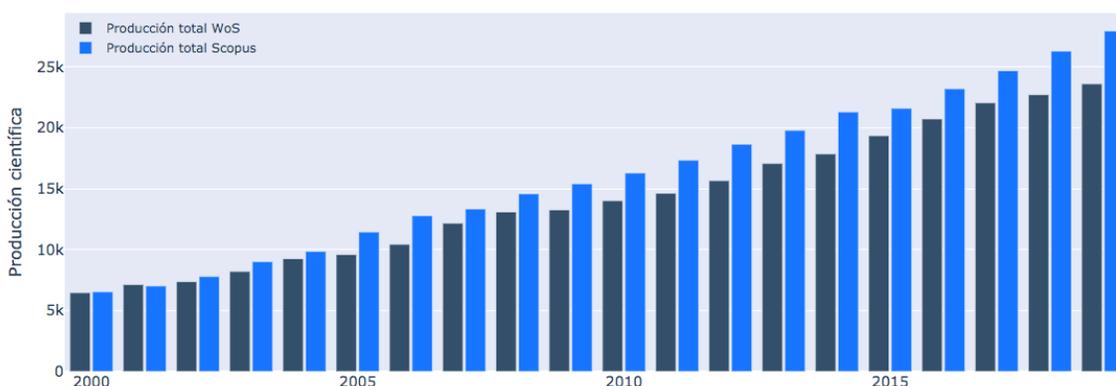


Figura 10. Evolución de la producción de México en *Web of Science* y Scopus, 2000-2019.

Específicamente, las tasas de crecimiento de la producción científica en WoS crecieron en promedio seis veces más que la economía y el total de la población durante el período, como se observa en la figura 11. En el año 2007 experimentó el mayor crecimiento de la producción total en WoS (16%). Este crecimiento de la producción científica mexicana puede estar relacionado con las políticas de incentivos del SNI para motivar a los investigadores a publicar.

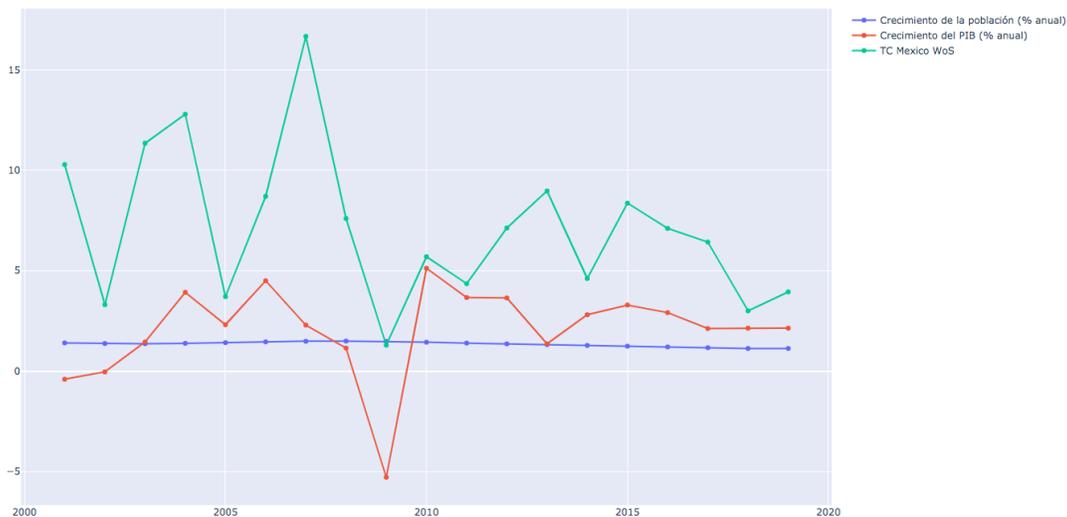


Figura 11. Comparación de la dinámica de las tasas de crecimiento de la producción científica, el producto interno bruto y la población en México durante 2000-2019.

La figura 12 muestra la evolución de la producción científica mexicana según el tipo de producción científica en el WoS. México, ha cuadruplicado su producción en revistas con factor de impacto durante los últimos 20 años. Con una tasa de crecimiento promedio en revistas con factor de impacto: 7.81%.

La producción en revista con factor de impacto ha aumentado cuatro veces durante el período, pasando de 4,938 documentos en el 2000 a 20,475 en el 2019, a una tasa promedio anual de crecimiento del 7.81%, alcanzando un volumen acumulado de 22,1575 documentos. Mientras que, el número de trabajos en revistas sin factor de impacto no presenta una tendencia importante de crecimiento, publicando un total de 1,531 trabajos en el año 2000 y 1016 documentos en el 2019 con una tasa de crecimiento promedio anual del 1.37% (figura 12).

Este incremento de la producción mexicana en revista de impacto puede estar relacionado con la consolidación en la década pasada de varias bases de datos útiles para el desarrollo de estudios cuantitativos, como Scopus, el Web de la Ciencia y Dimensions. Es evidente que, la competencia ha generado que estas bases crezcan en cobertura, y este crecimiento en cobertura se refleja de manera general en el crecimiento de la producción científica de cada uno de los países. Por lo que, otra hipótesis que se puede considerar es cómo este crecimiento se deriva de la cobertura de las bases y no de la propia actividad científica nacional.

Tipos de producción científica mexicana en WoS

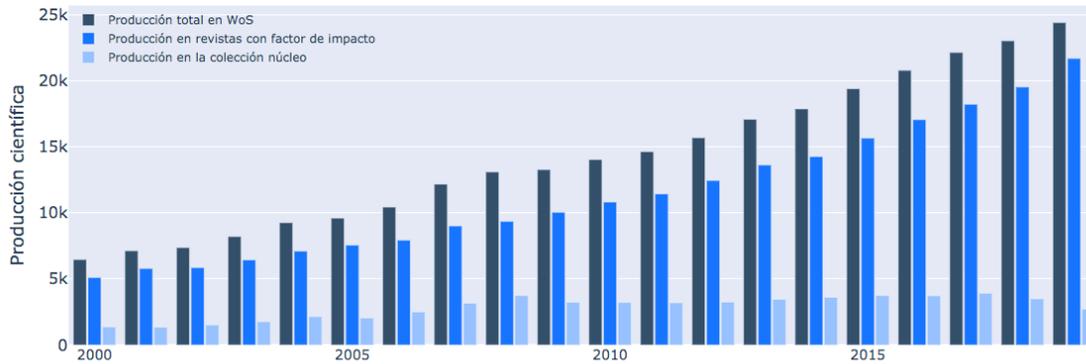


Figura 12. Dinámica de los tipos de producción científica mexicana en Web of Science, durante 2000-2019.

El país ha logrado triplicar la proporción de documentos publicados en revistas extranjeras (figura 13). Un total de 23,688 revistas científicas publicaron 285,981 trabajos mexicanos durante el período.

En la figura 13 se observa como la proporción de documentos publicados en revistas extranjeras se ha triplicado, pasando de 5,859 revistas en el 2000 a 19,758 revistas en 2019, mientras que el volumen de trabajos publicados en revistas se ha estancado.

Evolución de la producción científica mexicana en revistas nacionales e internacionales

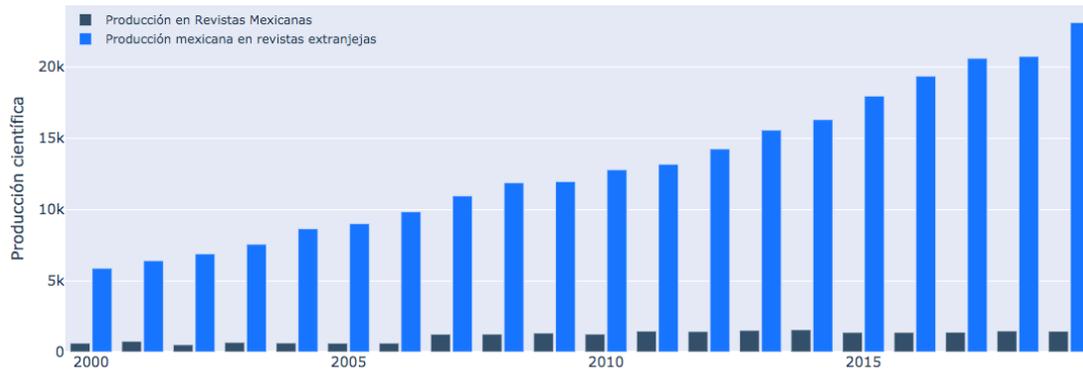


Figura 13. Dinámica del total de artículos mexicanos publicados en revistas nacionales y extranjeras indexadas en el WoS, durante el 2000-2019.

La figura 14 muestra la evolución de la producción científica mexicana y el número de miembros en el Sistema Nacional de Investigadores (SNI). El número de miembros del SNI está por encima que la producción durante el período. Existe un aumento en el incremento del número de SNI específicamente del año 2013-2017. El número de SNI se ha triplicado durante los últimos veinte años, pasando de 4,938 miembros en el 2000 a 28,236 en el año 2019 a una tasa promedio de crecimiento del 7.16% y creció a 5% durante los últimos cinco años (2015-2019).

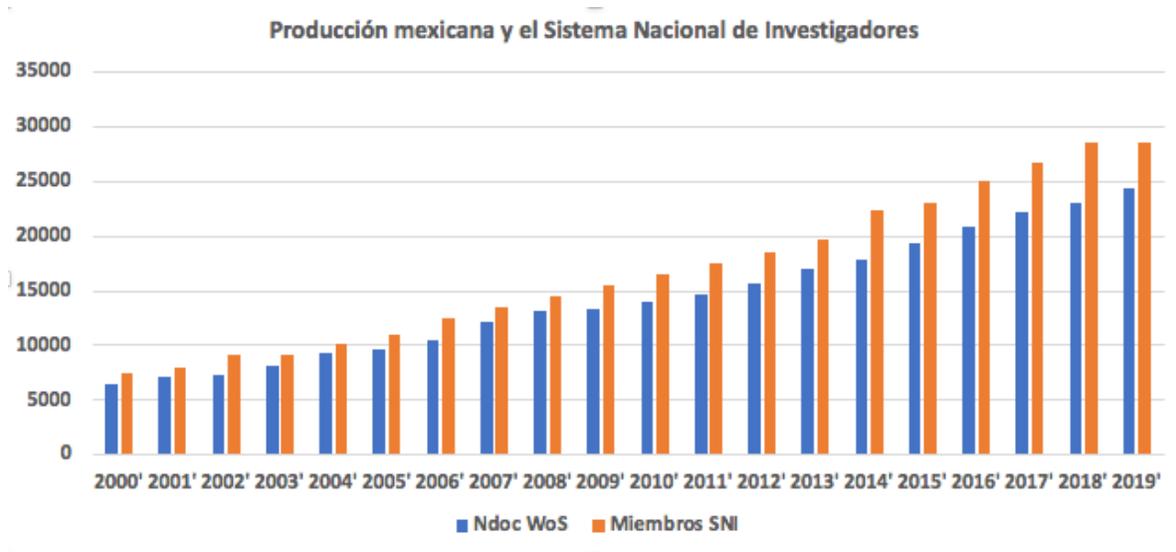


Figura 14. Evolución de miembros del Sistema Nacional de Investigadores y la producción total en WoS.

### 3.4.1. Índice de productividad nacional

En la Figura 14 observamos que la producción de artículos científicos y el número total de miembros del SNI han aumentado constantemente durante el período 2000-2019. Sin embargo, las diferentes tendencias en sus tasas de cambio han impactado negativamente el indicador nacional de productividad científica (NSP, producción entre miembros del SNI). La evolución del indicador nacional de productividad científica está representado en la figura 15 para los tipos de producción en el WoS: la producción total (Ndoc WoS); la publicada en revistas con factor de impacto y la producción en la colección núcleo (NSP WoS, NSP WoS-JIF y NSP PC Col Núcleo).

El NSP WoS osciló alrededor de un valor medio de 0.92 durante 2000-2019, pero, ha disminuido a un valor mínimo de 0.88. En la Figura 15, los indicadores NSP muestran una tendencia creciente durante la primera parte del período 2000-2019, pero esta tendencia creciente se detiene en la última parte del período. De hecho, el indicador NSP WoS-JIF decrece a partir de 2004, alcanzando su valor mínimo en 2018.

Los valores del NSP PC Col Núcleo son muy pequeños lo que demuestra cómo los investigadores y autores mexicanos son más productivos al publicar sus resultados en revistas con factor de impacto.

Esta disminución del índice de productividad puede estar relacionado con el debate internacional en relación con las políticas de evaluación. Hay muchos investigadores abogando por métricas responsables para la evaluación de la investigación, muchos están pidiendo el abandono de indicadores como el factor de impacto o el índice H de los ejercicios de evaluación, otros piden la inclusión de medidas de popularidad en la web o Altmétricas, otros se inclinan por la eliminación de todo paga en un proceso editorial. En fin, existen diversos elementos influyendo en este escenario, e incluso en el crecimiento del ecosistema de revistas nacionales.

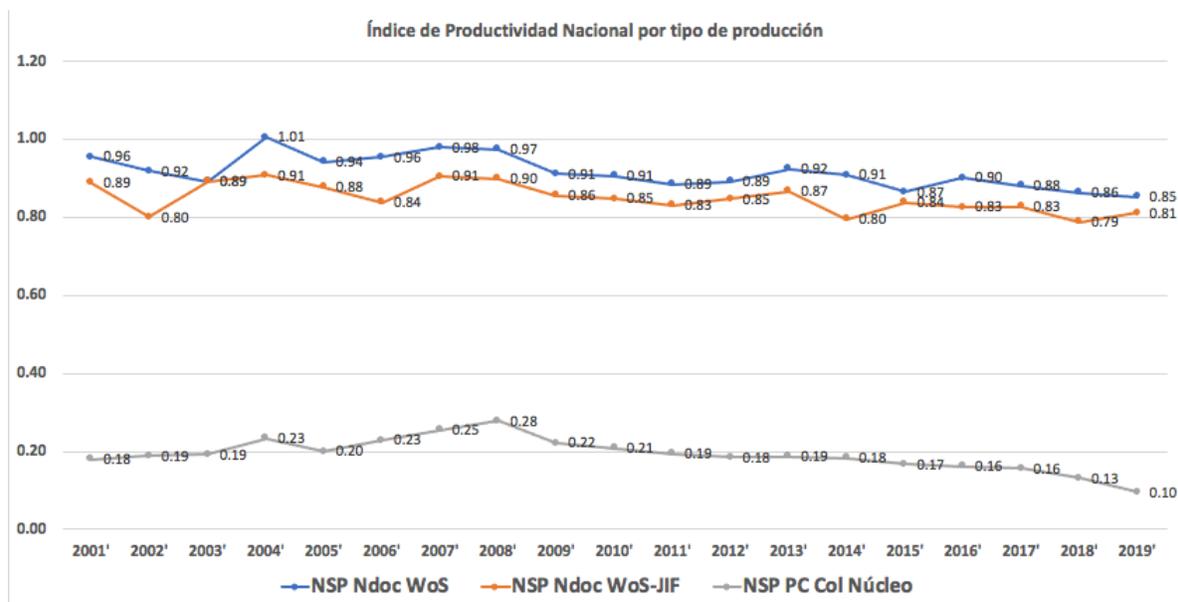


Figura 15. Evolución de la productividad nacional mexicana por tipo de producción en WoS, 2000-2019.

### 3.4.2. México en el contexto Latinoamericano

Para el análisis comparativo de la producción científica mexicana se seleccionaron los 10 países latinoamericanos con mayor número de documentos publicados en el WoS en el período 2000-2019. En este caso la producción de conocimiento en América Latina está encabezada por países como Brasil, México, Argentina y Chile.

Sin embargo, la evolución de la producción científica en relación con el número de habitantes en el período 2000-2019 cambió, como lo muestra la figura 16. En la región, Chile es el país que produce mayor cantidad de investigaciones científica en el WoS por cada 10,000 habitantes durante el período. Le sigue Uruguay, que a partir del año 2009 superó a Brasil y Argentina. México ocupa la séptima posición y

desde el año 2016, el país produce la misma cantidad de trabajos en esta base de datos por cada 10,000 habitantes que Colombia.

La gráfica 16 muestra que, la producción científica en WoS de los diez países latinoamericanos seleccionados crece a mayor velocidad anualmente que, la población.

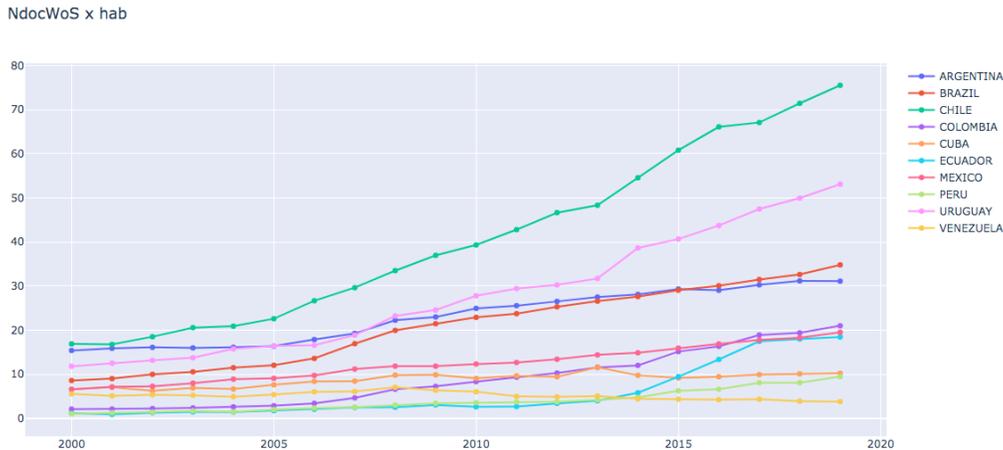


Figura 16. Evolución del volumen de artículos en WoS por cada 10,000 habitantes para diez países Latinoamericanos más productivos.

En la figura 17 comparamos los datos de los diez países latinoamericanos desde dos perspectivas: la fracción de la producción científica que estos países tienen en la producción mundial y su impacto normalizado.

Perú, Uruguay, Ecuador y Chile, con un bajo por ciento de su contribución a la producción mundial, alcanzaron un impacto por encima de la media mundial. Sin embargo, Brasil y México tienen un porcentaje mayor de sus producciones en la producción mundial, pero todavía el impacto de sus investigaciones no alcanzan la media mundial.

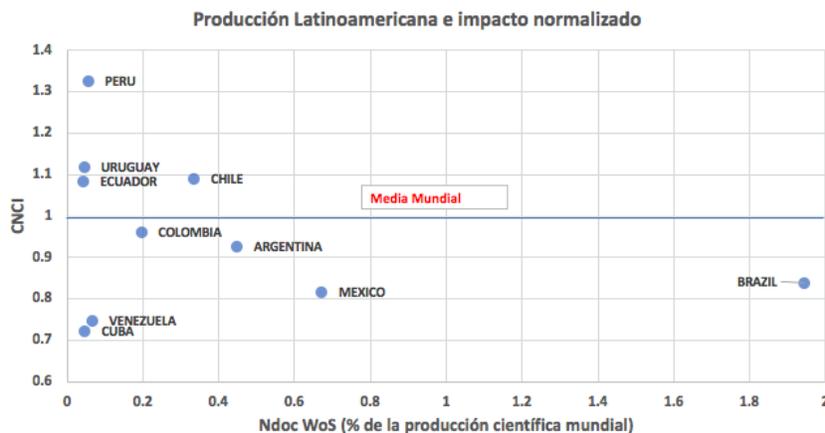


Figura 17. Comparación entre la producción por cada país de Latinoamerica como porcentaje de la producción mundial y el impacto normalizado (2000-2019)

El siguiente diagrama de dispersión relaciona la producción total, el liderazgo en la investigación y el por ciento de la producción científica de excelencia. Se identifican un conjunto de países como: Perú y Ecuador donde con una baja producción científica, sus investigaciones de excelencia duplican la media mundial, pero el por ciento de liderazgo nacional está por debajo del 52%.

Países como: Chile y Colombia con una baja producción tienen publicaciones de excelencia por encima de la media mundial. El 57% y 58% de estas investigaciones de excelencia cuentan con un líder del país de origen.

Por su parte, Brasil, México y Argentina son los mayores productores de la región, y más del 65% de sus investigaciones de excelencia son lideradas por autores nacionales. Sin embargo, todavía el impacto de la investigación de excelencia no alcanza la media mundial.

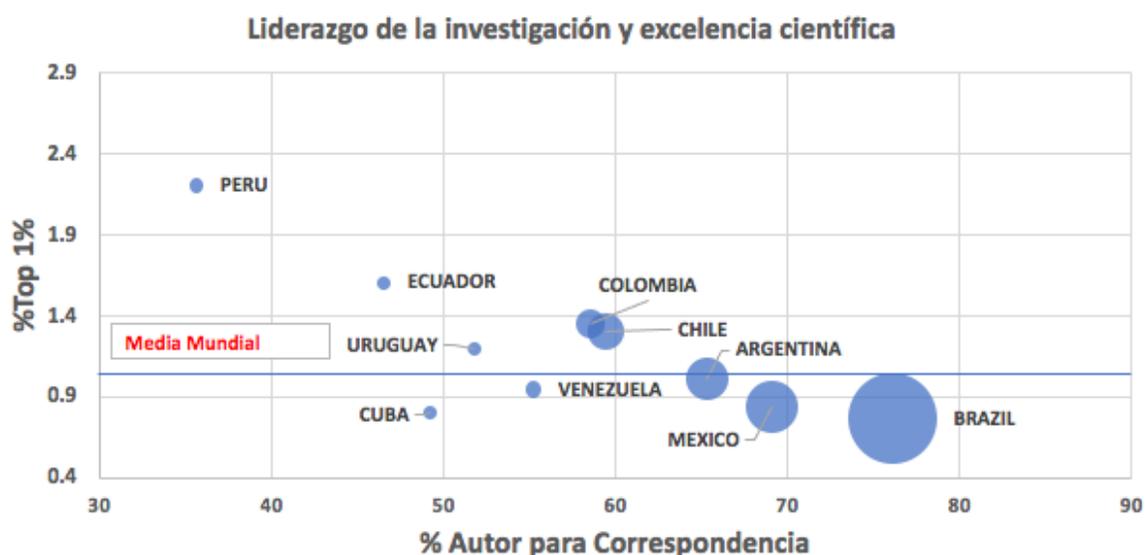


Figura 18. Relación entre la producción científica de excelencia y el liderazgo latinoamericano en la investigación, durante 2000-2019. El tamaño de las bolas representa el número de artículos en el período.

Algunas de las cuestiones reveladas por el análisis bidimensional se comprenden mejor desde una perspectiva multidimensional. En el siguiente mapa, obtenido bajo el enfoque de la inteligencia artificial y basado en técnicas neurocomputacionales, servirá para obtener una caracterización multidimensional de los perfiles cientiométricos de los diez países latinoamericanos más productivos en WoS.

Ahora se procede a realizar el análisis y visualización del perfil cientiométrico multidimensional de los diez países latinoamericanos más productivos, de acuerdo a siete indicadores: Prod x 10,000 Hab, CNCI, % Top 1%, % Top 10%, % CI, % Q1y %ACorr. En lugar de obtener siete rankings diferentes para cada uno de estos indicadores, o analizar el conjunto de datos desde la perspectiva de dos indicadores mediante diagramas de dispersión. Aquí nos enfrentamos al problema de comparar datos multidimensionales teniendo en cuenta los siete indicadores cientiométricos simultáneamente, aplicando técnicas de *machine learning*.

Para ello, primero representamos a cada país como un punto en un espacio euclidiano de siete dimensiones, cuyas coordenadas vienen dadas por los valores de los siete indicadores. Luego, usamos ViBlioSOM 2.0 para lidiar automáticamente con el problema de comparación y visualización. La técnica de visualización arroja ocho mapas (un mapa de clustering y 7 mapas de componentes) que muestran la comparación del rendimiento y la agrupación realizada por la red neuronal.

Estos mapas se dibujan sobre una cuadrícula hexagonal, donde cada hexágono representa una neurona. En este caso utilizamos una cuadrícula de 10 x 10 neuronas (Figura 19). Una vez que se entrena la red neuronal mediante un procedimiento adaptativo y no supervisado, se traza cada país en el mapa aplicando una proyección no lineal desde el espacio 7D a la cuadrícula neuronal.

Dado que estamos usando una medida de similitud que considera los siete indicadores cientimétricos, los países cercanos sobre la cuadrícula comparten perfiles cientimétricos multidimensionales similares.

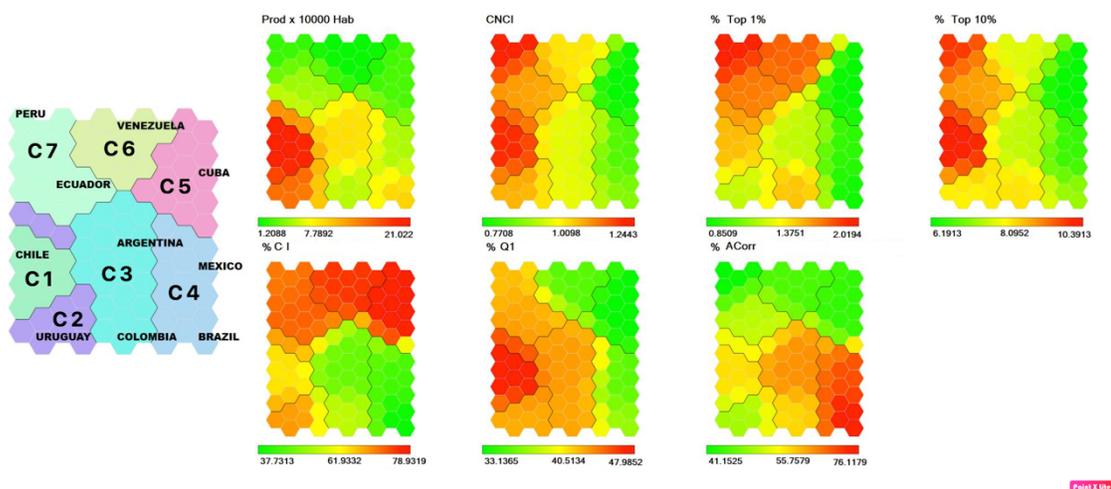


Figura 19. Perfil cientimétrico multidimensional de los 10 países latinoamericanos más productivos en WoS , durante el período 2017-2019.

En la figura 19 observamos ocho mapas, el primero de conglomerados o *clustering* y le siguen siete mapas de componentes. El mapa de *clustering* exhibe varias regiones (clusters) que se distinguen por colores y el grupo de países que pertenecen a cada región. Los países latinoamericanos de cada conglomerado tienen perfiles cientimétricos similares, de acuerdo con los siete indicadores seleccionados.

Los mapas de componentes, es otra de las visualizaciones que arroja ViBlioSOM 2.0. Existe un mapa para cada uno de los siete indicadores que componen el perfil cientimétrico multidimensional de los países latinoamericanos. Estos mapas son fundamentales para la interpretación de resultados. Mediante una escala cromática muestran la distribución del mapa de los valores de los indicadores. En cada mapa de componentes el valor máximo del indicador corresponde al rojo más oscuro, el valor medio está asociado al tono amarillo y el valor mínimo al verde. El amarillo intenso se asocia a valores cercanos a la media de cada indicador. Por lo tanto, los

mapas de componentes nos permiten leer el significado de la distribución de datos en la red neuronal.

El mapa de clustering de la Figura 19, exhibe que la red neuronal está dividida en siete grupos o cluster (C1,..., C7). Chile en el cluster C1 presenta un perfil cientométrico multidimensional atípico (altos porcentaje en todos los indicadores) en comparación con el resto de los países. Chile con mayor por ciento de artículos por cada 10,000 habitantes (21%) en la región y su perfil cientométrico multidimensional se caracteriza por un impacto superior a la media mundial; el por ciento de investigaciones de excelencia y alto desempeño superan la media mundial y más del 60% de sus publicaciones se realizan en colaboración internacional. Es un país donde más del 47% de las investigaciones alcanzan una alta visibilidad esperada y el 55% de sus publicaciones son lideradas por un autor nacional.

Las investigaciones de los países que pertenecen a los cluster C2, C3, C6 y C7, tienen un impacto normalizado por encima de la media mundial. Además, los perfiles cientométricos de estos países tienen altos porcentajes de su producción en el top 1% y 10 % de lo más citado a nivel mundial. En promedio más del 68% de sus investigaciones se publican en colaboración internacional. Sin embargo, tienen un bajo por ciento de publicaciones por cada 10,000 habitantes; en promedio el 45% de sus trabajos se publican en revistas de primer cuartil y todavía, es bajo el por ciento de investigaciones con liderazgo nacional.

México y Brasil, con perfiles cientométricos multidimensionales similares se ubican en el cuarto cluster (C4). Sus perfiles muestran valores por debajo de los valores promedio en casi todos los indicadores. Sin embargo, ambos países son los únicos de la región con mayor porcentaje de liderazgo nacional en sus investigaciones.

Cuba se ubica en el cluster (C5). Su perfil cientométrico multidimensional se caracteriza por una baja producción por habitantes, el impacto de sus trabajo no alcanzan la media mundial; tiene un bajo por ciento de su producción en el top1% y top10% de lo más citado a nivel mundial. Asimismo, es baja la visibilidad esperada y el liderazgo. Sin embargo, es el país con mayor porcentaje de publicaciones en colaboración internacional.

### 3.4.3. México en el contexto mundial

Una vez comparado el perfil cientométrico multidimensional de México y los países latinoamericanos más productivos en WoS, aplicamos nuevamente ViBlioSOM 2.0 para identificar los países con perfiles económico-demográficos similares a México y luego comparamos su desempeño cientométrico.

Se seleccionó una muestra de 108 países en la *Web of Science* para obtener cuatro indicadores económico-demográficos (figura 20): Población en 2019 (Pop 2019), Índice de Competitividad Global (GCI), Índice de Complejidad Económica (ECI) y el producto interno bruto per cápita (PIB per cápita 2019).

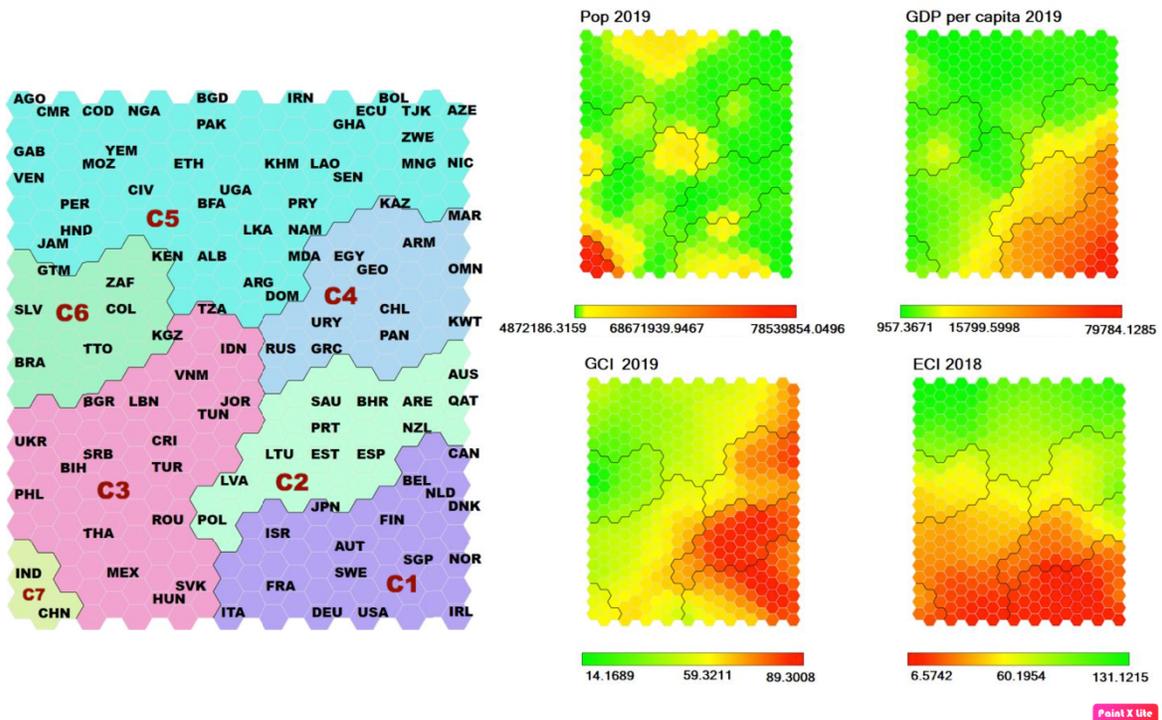


Figura 20. Comparación de 108 países utilizando indicadores económico-demográficos. Los códigos de país utilizan el estándar internacional ISO 3166-1 alpha-3.

La red neuronal agrupó la muestra de 108 países en siete conglomerados que comparten las características económico-demográficas más similares (Figura 20).

En la figura 20 los cuatro mapas de componentes correspondientes a cada indicador muestran los países con las economías más complejas y los valores más altos de PIB per cápita. Se agrupan en los cluster C1, C2 y C4. En el grupo C7, China e India aparecen como valores atípicos debido a su alto volumen de población.

Dieciocho países pertenecen al clúster C3, compartiendo valores altos del Índice de Complejidad Económica y valores medios en los demás indicadores. Este grupo incluye dos países latinoamericanos (Costa Rica y México) y ocho países europeos (Bosnia y Herzegovina, Bulgaria, Turquía, Rumania, Serbia, Hungría, Ucrania, Eslovaquia), cuatro países asiáticos (Indonesia, Filipinas, Tailandia y Vietnam), dos africanos (Túnez y Tanzania) y dos países de Oriente Medio (Jordania y Líbano). Considerando que estos 18 países son los más parecidos a México en cuanto a los indicadores económicos-demográficos, procedemos a comparar sus perfiles de desempeño cuantitativo multidimensional utilizando la red neuronal (Figura 21).

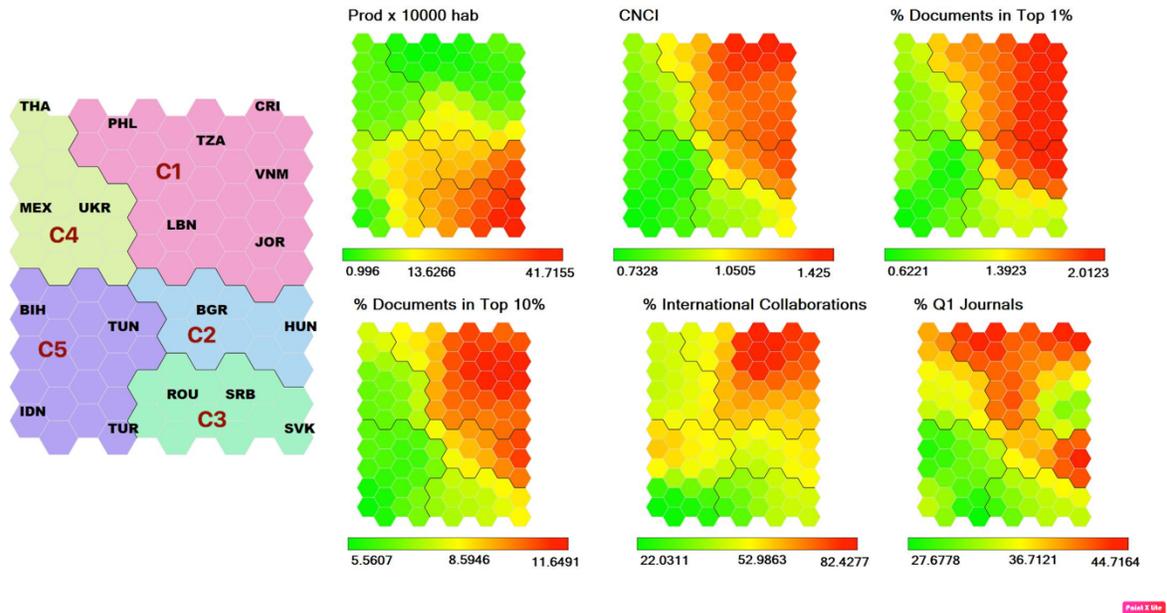


Figura 21. Perfiles cientimétricos multidimensionales de países con perfiles económicos-demográficos más similares a México (2017-2019).

De los cinco clusters identificados por la red neuronal, los países con mejores perfiles de desempeño se encuentran en los grupos C1 y C2. Los países del cluster C2 y C3 se destacan como los más productivos y C1 tienen el puntaje de colaboración internacional más alto. México, Ucrania y Tailandia pertenecen al C4. Este clúster junto con C3 mostró los valores más bajos en los indicadores de productividad e impacto. Sin embargo, C4 tiene un mejor porcentaje de artículos en el primer trimestre, mientras que C3 tiene un mejor nivel de productividad.

Centrándonos en el desempeño cientimétrico de México como ejemplo, los mapas de la figura 21 muestran que, el país está superado en productividad e impacto por los 11 países que pertenecen a los conglomerados C1, C2 y C3. Por ejemplo, el 0,94% de la producción es de excelencia frente a un máximo del 2% marcado por algunos países en los conglomerados C1 y C2. El 6,69% de sus documentos son de alto desempeño frente a un máximo de 11,6 para algunos países en estos grupos. De igual forma, en impacto normalizado, México tiene 0.84 mientras que los países con mejor puntaje alcanzan 1.4. En cuanto al indicador de productividad, México tiene 5 artículos por 10,000 habitantes mientras que el puntaje mayor en el clúster C3 es 42 y Tanzania tiene un mínimo de 0.8.

### 3.5. Mapeo de los perfiles de desempeño cientimétrico de las instituciones mexicanas.

#### 3.5.1. Colaboración internacional

En los últimos 20 años México ha colaborado con un total de 8003 instituciones de 144 países. En el treemap (Figura 22) están representadas las 50 instituciones extranjeras que tienen mayor producción científica en colaboración con México. Los

investigadores mexicanos han logrado desarrollar proyectos conjuntos y relaciones de colaboración con instituciones altamente reconocidas en la comunidad científica internacional: Centre National de la Recherche Scientifique (CNRS); el Consejo Superior de Investigaciones Científicas (CSIC), Harvard University (HU), University of California System (UC).

En este grupo, los trabajos publicados en colaboración con el Lomonosov Moscow State University (LMSU), Purdue University system (PUS), University of Arizona (UA), Imperial College London (ICL), University of Illinois System (UIC), Texas A&M University College Station (Texas A&M) y el National de la Recherche Scientifique (CNRS) tienen el mayor impacto promedio por cada investigación, siendo cinco veces mayor a la media mundial.

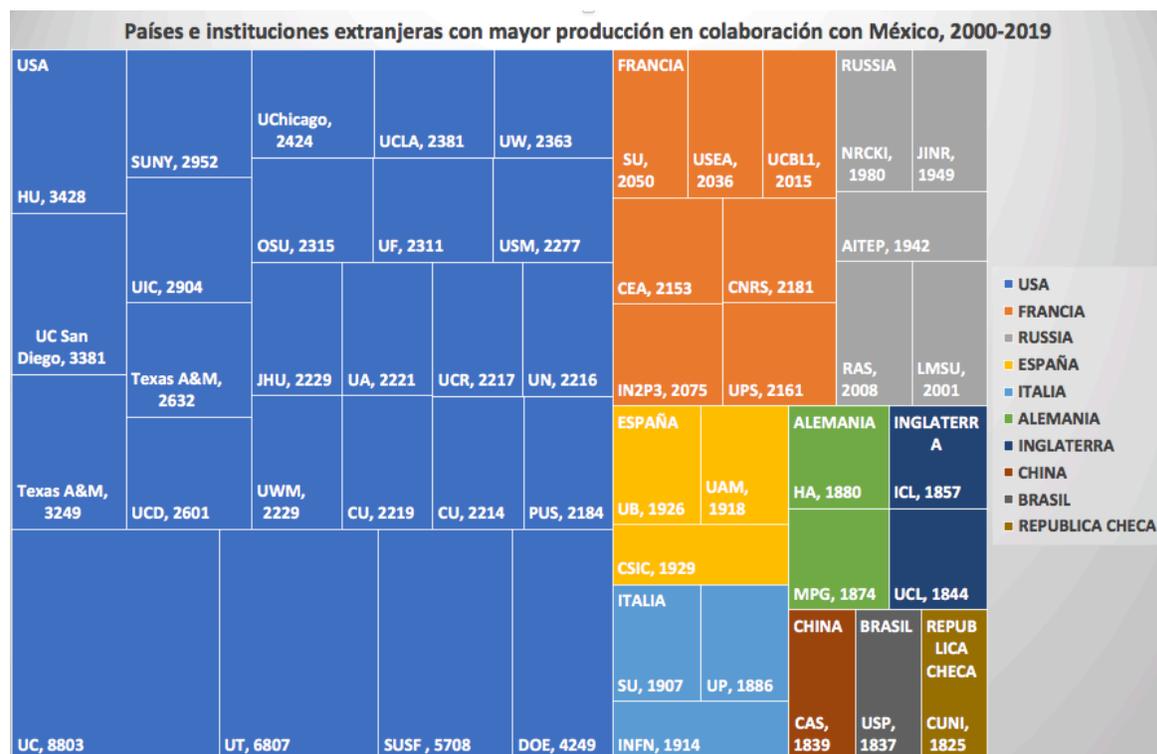


Figura 22. Instituciones y países con mayor colaboración científica con México en el WoS, 2000-2019. Los acrónimos de las instituciones se pueden consultar en el Anexo 2.

El 21% (59,156) de los trabajos mexicanos en colaboración con instituciones extranjeras tienen un impacto normalizado por artículo entre 10 y 100 veces mayor a la media mundial (Figura 23). Este conjunto de investigaciones en su mayoría pertenecen al área de la medicina clínica, se publicaron en revistas de primer cuartil, son investigaciones que por el volumen de citas que recibieron son artículos altamente citados y están en el 1% y 10% de lo más citado a nivel mundial.

Ejemplo de ello es el caso de los artículos con el National Acoustic Laboratories, The Morgridge Institute for Research, Inc., Kenema Government Hospital, Ministry of Health & Sanitation Sierra Leone, Vector Control Research Center, Landstuhl Regional Medical Center, Novartis Turkey, NHS Blood & Transplant (NHSBT), Beijing Neurosurgical Institute, National Institute of Occupational Health & Poison Control, Chinese Center for Disease Control & Prevention y University of Port Harcourt.



Figura 23. Impacto de la investigación mexicana con instituciones extranjeras, 2000-2019.

### 3.5.2. Producción e impacto de las instituciones más productivas

En el WoS se registran un total de 89 instituciones mexicanas. Nuestro estudio analiza 60 instituciones mexicanas con más de 40 artículos publicados por año en el período 2015-2019. El período de estudio se redujo con el objetivo de obtener un análisis más actualizado a nivel meso.

A continuación, descendiendo a nivel meso, realizamos un análisis comparativo de la dinámica de la producción científica en las 60 instituciones mexicanas con mayor rendimiento en WoS en el período 2015-2019. Para esta comparación, primero mostramos diagramas de dispersión de varios pares de indicadores.

La Universidad Nacional Autónoma de México (UNAM) se destaca por tener el mayor número de miembros en el SNI y volumen de producción científica durante el período estudiado. Produciendo 25,512 artículos, que representan el 27,68% de la producción nacional. Su producción media anual en este período fue de 5,102

artículos y el impacto normalizado de sus investigaciones es igual a 0.8. Para obtener un índice de productividad institucional (ISP) igual a 5,60. Al ser un valor atípico por su puntaje de producción, lo excluimos de algunos de los gráficos de este artículo (Figura 24 y 25) para obtener una mejor imagen del comportamiento de las otras instituciones.

Solo el Instituto Politécnico Nacional (IPN; 10,184) y el Centro de Investigaciones y Estudios Avanzados (CINVESTAV; 6,188), tuvieron una producción promedio anual superior a 1000 artículos. Estas instituciones concentraron 11,05 y 6,71% de la producción científica mexicana y obtuvieron un índice de productividad institucional por encima de la media (8,55 y 8,48). Estas instituciones fueron excluidas de la figura 24 por sus altos valores en la producción y el impacto normalizado.

La Benemérita Universidad Autónoma de Puebla, Universidad Autónoma Metropolitana - Mexico y el Tecnológico de Monterrey con más de 3500 artículos sus trabajo logran alcanzar un impacto normalizado por encima de la media mundial (BUAP CNCI=1,23, UAM CNCI=1,12 y ITESM CNCI=1,11).

Hubo diez instituciones con menos de 2000 artículos en el período alcanzaron un impacto normalizado hasta cuatro veces mayor a la media mundial. Ejmplo: Instituto Nacional de Psiquiatria Ramón de la Fuente Muñiz (INPRFR CNCI= 3,99); Instituto Nacional de Salud Pública INSP (CNCI=3.33); Instituto Nacional de Cancerología (INCAN CNCI= 2.05).

El resto de las instituciones tienen baja producción y todavía sus trabajos no logran un impacto internacional por encima de la media mundial.

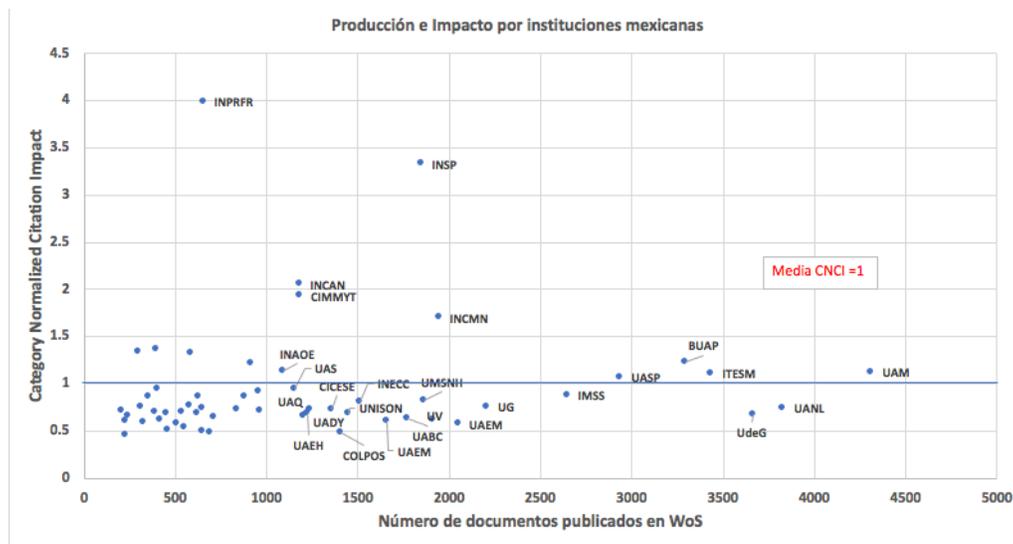


Figura 24. Producción e impacto de las 60 instituciones mexicanas más productivas en WoS, durante el 2015-2019.

La UNAM, CINVESTAV e IPN no aparecen en la figura 24 porque muestran altos valores en el número de artículos publicados. UNAM (NdocWoS=25,512; CNCI=0,84); CINVESTAV (NdocWoS=6,188; CNCI=1,14) y el IPN(NdocWoS=10,184; CNCI=0,77).

El indicador NSP tiene un valor promedio de 0,92 durante el período 2015-2019, y las 60 instituciones más productivas tienen un valor promedio de ISP de 5,79. Entonces estas 60 instituciones tienen un aporte relevante a la productividad científica nacional (Figura 25).

Un conjunto de veintinueve instituciones tuvieron valores de ISP superiores a 5, con una producción científica menor a 1500 trabajos en el período, entre ellos: Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT; 12,44), Instituto Nacional de Cancerología (INCAN; 12,28), Instituto Nacional de Medicina Genómica (INMEGEN; 12,25), Instituto de Ecología - Mexico (INECC; 11,91) y Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubiran (INCMN; 10,92) (Figura 25).

Tres instituciones en la muestra seleccionada, produjeron entre 5 y 8 artículos por año (visibles en WoS): Instituto Mexicano del Seguro Social (IMSS; 8,29), Tecnológico de Monterrey (ITESM; 6,76) y la Universidad Autónoma de San Luis Potosí (UASP; 5,93). El resto de las instituciones produjeron más de dos artículos por año (Figura 25).

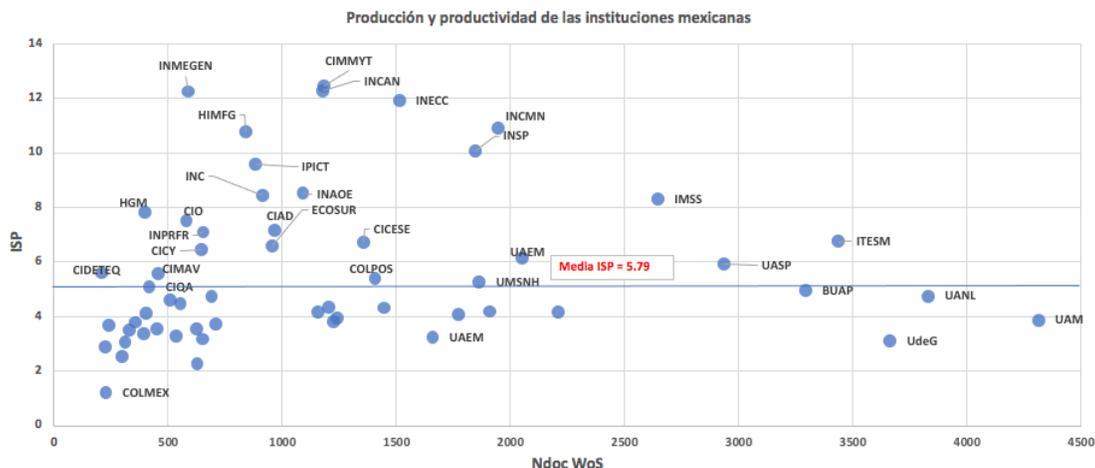


Figura 25. Producción y productividad (Institutional Scientific Productivity) de las 60 instituciones mexicanas más productivas en WoS, durante el 2015-2019.

La UNAM, CINVESTAV e IPN no se muestran en la figura 25 porque muestran altos valores en el número de artículos publicados. UNAM (NdocWoS=25,512; ISP=5,60); el IPN (NdocWoS=10,184; ISP=8,55) y CINVESTAV (NdocWoS=6,188; ISP=8,48).

El Instituto Nacional de Salud Pública y el Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz con una producción científica baja sus trabajos alcanzan un impacto cuatro veces mayor a la media mundial (INPRFR; 3,99 y INSP 3,33). Además, en promedio sus investigadores producen más de siete artículos por año en el WoS (Figura 26).

Instituciones poco productivas generan investigaciones que logran obtener un impacto normalizado por encima de la media mundial (Figura 26). Ejemplo: Instituto Nacional de Cancerología (INCAN; 2,05), Centro Internacional de Mejoramiento de Maíz y Trigo (CIMMYT; 1,93), Instituto Nacional de Ciencias Médicas y Nutrición

Salvador Zubiran (INCMN; 1,70) y Instituto Nacional de Medicina Genómica (INMEGEN; 1,32). Nueve instituciones con una producción entre 2 y 8 artículos por investigador en el período logran un impacto normalizado por encima de la media. Entre ellas: Hospital General de México (HGM; 1,36), Instituto Nacional de Cardiología (INC; 1,22), Instituto Nacional de Astrofísica, Óptica y Electrónica (INAOE; 1,13) y el Centro de Investigaciones y Estudios Avanzados (CINVESTAV; 1,14).

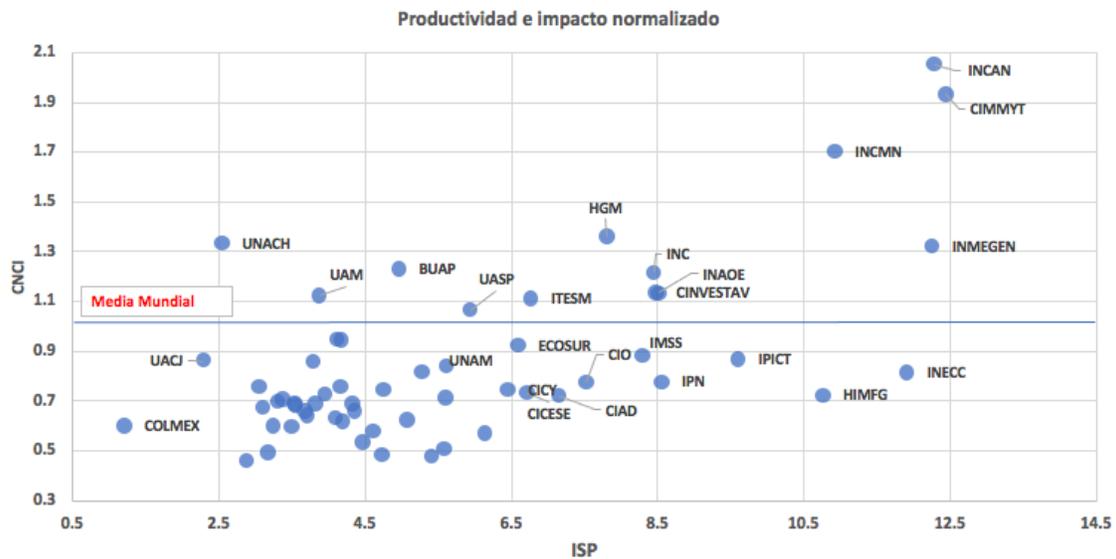


Figura 26. Impacto normalizado (*Category Normalized Citation Impact*) versus la productividad institucional (ISP) durante 2015-2019.

Instituto Nacional de Salud Pública e Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz no se incluyen en la gráfica 26 porque tienen los valores más alto para ambos indicadores. Ver anexo 6

### 3.5.3. Impacto esperado y alto desempeño de las instituciones mexicanas

Las instituciones con un mayor por ciento de publicaciones en revistas de primer cuartil reciben un cúmulo mayor de citas, ubicándose en el 10% de lo más citado en su campo, año y tipo de documento (figura 27). Al contrario de lo que cabría esperar, el gráfico de dispersión de % Top10% versus %Q1 muestra un grado bajo de correlación entre los indicadores (coeficiente de correlación  $R^2 = 0.5379$ ).

Sin embargo, once instituciones tienen puntajes por encima de los valores medios de ambos indicadores. INPRFR muestra el comportamiento más atípico: en %Q1 (41,47) y con la puntuación más alta en % Top10% (14,55). Las once instituciones ubicadas en el cuadrante superior derecho sugiere que, el logro de alto desempeño está ligado a la publicación en revistas de alta visibilidad (figura 27).

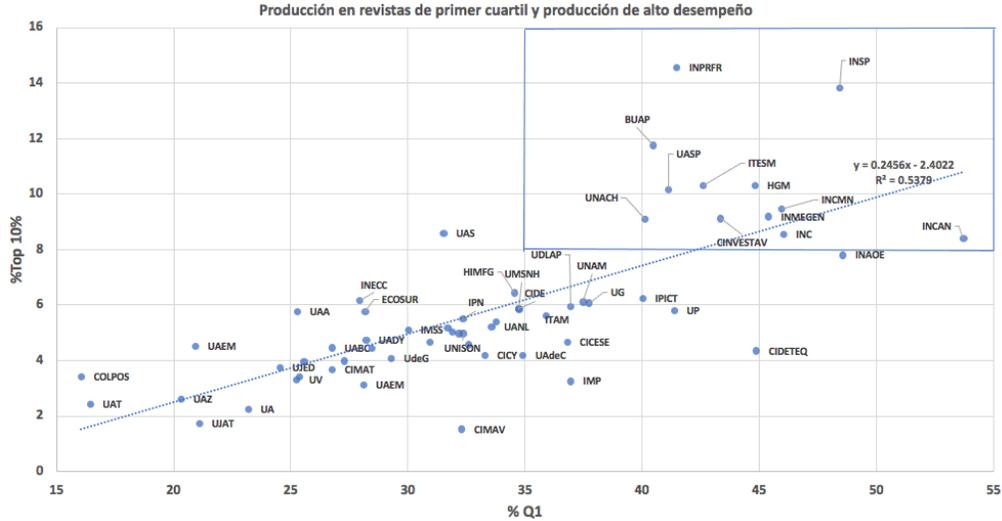


Figura 27. Producción de alto desempeño y visibilidad esperada en 60 instituciones mexicanas más productivas en WoS, durante 2015-2019.

El Centro Internacional de Mejoramiento del Maíz y Trigo no parece en la gráfica 27 por sus altos valores en ambos indicadores: %Q1= 60.70 y %Top 1%= 23.01

### 3.5.4. Comparación de los perfiles cienciométrico multidimensionales de las instituciones mexicanas

Esta sección aborda desde el enfoque de la inteligencia computacional, basada en técnicas de mapas auto-organizados o SOM (Self-Organizing Map), una caracterización multidimensional. El mapa de la figura 28 muestra el mapeo multidimensional sobre los perfiles cienciométricos de las 60 instituciones mexicanas más productivas, de acuerdo con los indicadores: Índice de productividad institucional (ISP), *Category Normalized Citation Impact* (CNCI), % documentos en Q1 (%Q1), % documentos en el 1% de lo más citado (%Top1%), % documentos en el 10% de lo más citado (%Top10%), % documentos en colaboración internacional (%CI) y liderazgo (%Acorr).

En este caso se utilizó una rejilla de 20x20 neuronas y se realizaron 5000 iteraciones para entrenar la red neuronal en ViBlioSOM 2.0. El mapa de conglomerados de la figura 28 se observan diez clusters o regiones (C1,..., C10) y se distinguen por colores tres tipos de instituciones. Las instituciones que quedaron dentro de una misma región ofrecen perfiles cienciométricos multidimensionales similares, de acuerdo con los siete indicadores seleccionados.

Señalamos el caso del Centro Internacional de Mejoramiento del Maíz y Trigo; Instituto Nacional de Cancerología; Instituto Nacional de Salud Pública y el Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz; los cuales no se incluyeron en el mapa de la figura 28 porque sus perfiles cienciométricos multidimensionales son atípicos, presentando altos valores en la mayoría de los indicadores (Anexo 6).

En la figura 28 los comportamientos atípicos son representados por una sola institución en un cluster como el caso de: Hospital General de México (HGM) y el Colegio de la Frontera Sur (ECOSUR), porque son instituciones que cumplen con un alto desempeño con respecto a los indicadores estudiados y el grupo de 60

instituciones. Las instituciones que pertenecen al sector salud tienen perfiles multidimensionales cuantitativos por encima del resto de los perfiles.

La zona superior del mapa de conglomerados donde se ubican los clusters C2, C3, C4 y C5 tienen valores altos en ISP, CNCI, %Q1, %Top 1% y %Top10%. Las instituciones que pertenecen a estas regiones sus perfiles cuantitativos multidimensionales se caracterizan por publicar más de 5 artículos por investigador, los trabajos logran un impacto por encima de la media mundial y más del 45% de sus investigaciones son publicadas en revistas de primer cuartil. Además, logran tener publicaciones de excelencia y alto desempeño por encima de la media mundial. Sin embargo, no todas tienen un alto porcentaje de publicaciones en colaboración internacional y no tienen un alto porcentaje de publicaciones con liderazgo institucional. Ejemplo: Benemérita Universidad Autónoma de Puebla (BUAP), Universidad Autónoma de San Luis Potosí (UASP), Tecnológico de Monterrey (ITESM) y el Centro de Investigaciones y Estudios Avanzados (CINVESTAV).

Los perfiles de las instituciones que se ubican en la región C8 se caracterizan por publicar en promedio el 40% de sus trabajos en revistas de primer cuartil, más del 45% de sus trabajos cuentan con líderes institucionales y aproximadamente 40% de sus trabajos se escriben en colaboración internacional. Sin embargo, producen alrededor de un artículo por investigador; las investigaciones no alcanzan un impacto por encima de la media mundial y los porcentaje de artículos de excelencia y alto desempeño no superan la media mundial.

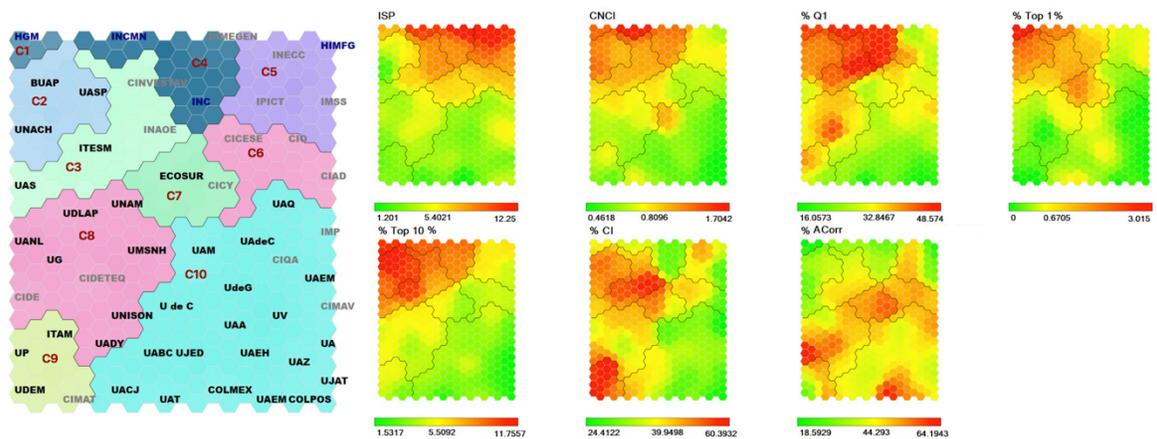


Figura 28. Perfil multiparamétrico de las 60 instituciones mexicanas más productivas en WoS, durante el período 2015-2019.

El Centro Internacional de Mejoramiento del Maíz y Trigo; Instituto Nacional de Cancerología; Instituto Nacional de Salud Pública y el Instituto Nacional de Psiquiatría Ramón de la Fuente Muñiz no parecen en la gráfica por sus altos valores en todos los indicadores. Ver los valores de las instituciones en el Anexo 6

**Color azul:** instituciones del sector salud; **Color negro:** universidades; **Color gris:** centros e instituciones de investigación

### 3.6. Mapeo de la evolución de las áreas de investigación de la ciencia mexicana.

#### 3.6.1. Participación de producción en cada área de investigación de la producción nacional.

En esta sección analizamos la producción científica publicada en revistas con factor de impacto en 22 áreas de investigación mexicanas bajo el esquema de clasificación ESI. En la figura 29 se compara la fracción de México en cada área; en relación a la producción nacional total en revistas de corriente principal.

Durante los 20 años estudiados (2000-2019), la producción mexicana en revistas con factor de impacto está dominada por *Clinical Medicine*, que representa más del 10% del total de publicaciones a nivel nacional (ver Figura 29) en los años 2000, 2005, 2010, 2015 y 2019. Le siguen: *Plants & Animal Science*, *Chemistry*, *Engineering*, *Environment/Ecology*, *Physic* y *Biology & Biochemistry*. En general, la mayor contribución en la ciencia de México proviene de las ciencias exactas.

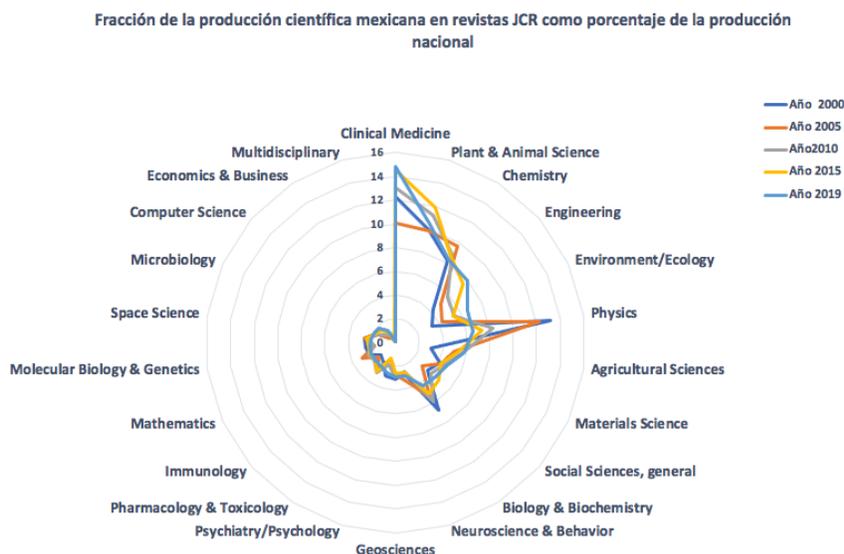


Figura 29. Participación de la producción académica mexicana en revistas con factor de impacto, 2000–2019.

#### 3.6.2. Índice de actividad científica

Sin embargo, al comparar la contribución de México con respecto al esfuerzo que el mundo le dedica a cada campo de conocimiento (Figura 30) el esfuerzo de la ciencia mexicana toma otro giro.

En la figura 30 evaluamos las ventajas comparativas de México e identificamos sus cambios a lo largo de los años 2000, 2005, 2010, 2015 y 2019 en relación al mundo. En nuestra investigación se calcula el *Revealed Comparative Advantage* (RCA) (Balassa, 1965, 1977) para conocer los cambios en las áreas de investigación mexicana. Específicamente, este indicador proviene del campo de la

economía pero ha sido aplicado en el campo de la cienciometría. Matemáticamente, este índice tiene la misma forma que el índice de actividad científica (AI). El puntaje de RCA de un campo temático es la proporción entre la fracción del total de México en una categoría y la fracción de esta categoría en el resto del mundo. Los puntajes RCA por campo revelan las fortalezas y debilidades de las capacidades de investigación en el país.

En el año 2019, *Space Science* es la disciplina en México, que contribuye más a la producción en revistas de factor de impacto en la participación mundial. Le siguen *Plants & Animal Science*, *Agriculture Sciences*, *Environment/Ecology* y *Physics* donde el esfuerzo del país es en promedio cinco veces mayor al mundo. Una observación interesante es que una gran contribución al total mundial no proviene necesariamente de las categorías donde México tiene mayor producción científica publicada en revistas con factor de impacto.

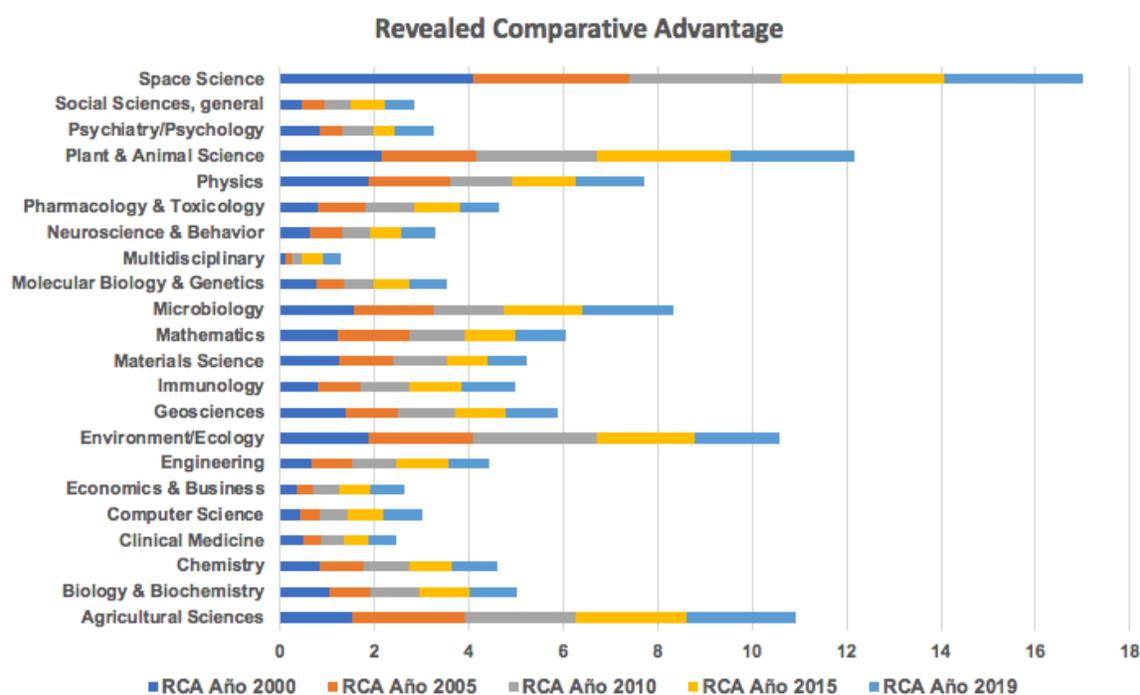


Figura 30. Revealed Comparative Advantage para las áreas de investigación mexicanas en 2000, 2005, 2010, 2015 y 2019.

En resumen, la figura 31 muestra dos importantes piezas de información. Una es que la mayoría de los campos han mejorado su RCA de 2005 a 2019. La otra es que, teniendo en cuenta el tamaño de las bolas, las categorías con menor participación en la producción científica nacional en revistas con factor de impacto han mejorado su contribución en la producción mundial de calidad. *Space Science*, *Plants & Animal Science*, *Agriculture Sciences*, *Environment/Ecology*, *Microbiology* y *Physics* tienen menor participación en la producción científica nacional, pero han aumentado su ventaja comparativa con respecto al mundo.

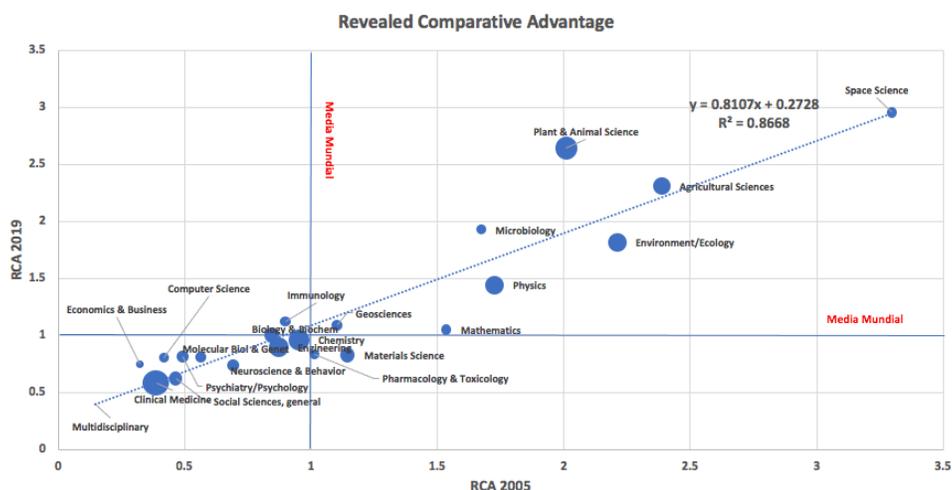


Figura 31. Revealed Comparative Advantage para México en 22 campos científicos en 2005 y 2019. El país es comparado con el resto del mundo. El tamaño de la bola está representado por la contribución de México en cada campo en la producción nacional en el 2019.

### 3.6.3. Mapeo neuro longitudinal de las áreas de investigación: 2017- 2019

En este apartado analizamos la evolución de 2017 a 2019, de los perfiles cuantitativos multidimensionales de las 22 áreas de investigación mexicanas, mediante un mapeo neuro longitudinal anual realizado con ViBlioSOM 2.0. Esta secuencia de mapas longitudinales neuronales, se presenta en la Figura 32, exhibiendo cambios en dos niveles: macro y meso. En el nivel macro, observamos variaciones de color en los mapas de calor, en el nivel meso, cambios en la organización de los conglomerados y desplazamientos de áreas individuales en el mapa de conglomerados.

#### *Análisis a nivel macro*

A nivel macro, observamos el progresivo predominio del color rojo, a medida que avanzamos de un año a otro, revelando una tendencia general de mejora de las 22 áreas de investigación mexicanas, de 2017 a 2019. Estas áreas muestran una mejora notable en los puntajes de excelencia (%Top1%), llegando al año 2019 con un mayor área en color rojo. En el mapa de componentes del CNCI y %Top10% observamos también algunas mejoras hacia 2018 y 2019, pero en los mapas de los otros indicadores no hay grandes cambios durante los tres años de análisis (Figura 32).

En la parte superior del mapa de clustering observamos que, se concentran las áreas de investigación mexicana con mejores perfiles cuantitativos multidimensionales. En el 2017 y 2018 solo había un área de humanidades por encima de la diagonal de color azul; la cual terminó en el año 2019 por debajo de la diagonal.

Excepcionalmente, las áreas que pertenecen a un clúster permanecen sin muchos cambios a lo largo de estos tres años. Las áreas de investigación mejor

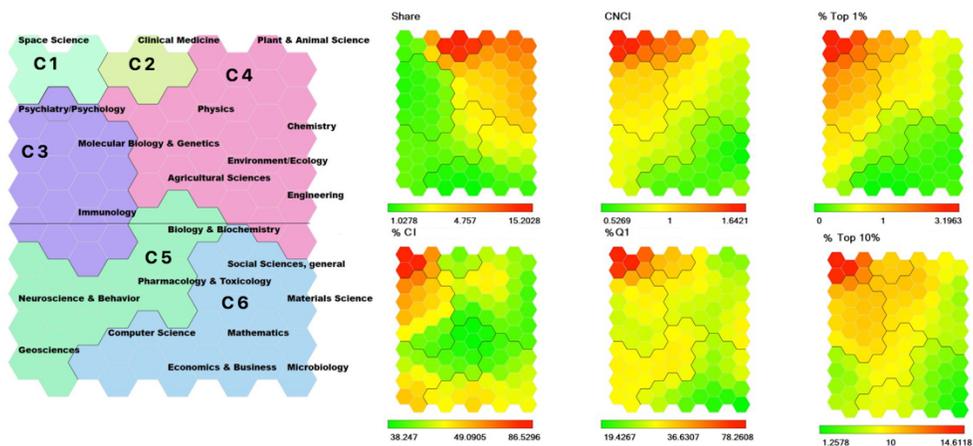
clasificadas *Space Science* (C1) y *Clinical Medicine* (C2) integraron un clúster durante los años 2017 y 2019. En el año 2018, cambió su composición del C1 al incluirse *Physic* y *Psychiatry/Psychology*. Durante el 2017 y 2018, *Plant & Animal Sciences* (C4), mantuvo un perfil similar al de otras áreas; cambiando su desempeño en el 2019, logró ocupar un solo cluster.

#### *Análisis de nivel meso*

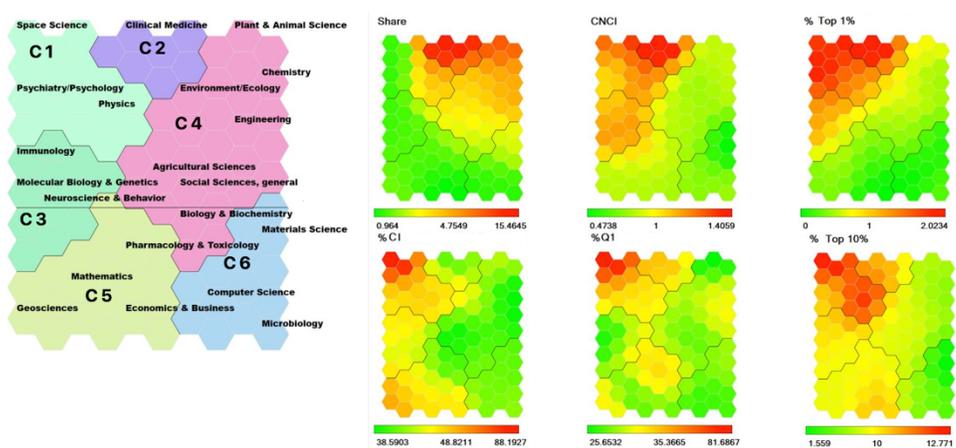
Los 6 grupos identificados en el año 2019 son el resultado de uniones o divisiones de clústeres durante años anteriores. La división de grupos se puede interpretar como una diferenciación de perfiles entre las áreas que forman un grupo, mientras que las uniones de grupos como una homogeneización de perfiles.

Los mapas de conglomerados muestran una variación durante 2017-2019. Observamos que las áreas de investigación entran o salen, de un año a otro, a la mayoría de los clusters. Por ejemplo, el grupo (C3) integrado por *Immunology*, *Molecular Biology & Genetics* y *Psychiatry/Psychology* en 2017, caracterizado por valores por encima del promedio (naranja) en cinco indicadores, en el año siguiente (2018) incluyó a *Neurosciences & Behavior*. Finalmente, en 2019, el clúster se dividió porque *Immunology*, *Molecular Biology & Genetics* y *Neurosciences & Behavior* se integraron al cluster número seis(C6). En este caso el clúster original conservó su perfil (C3), pero *Physic* y *Environment Ecology* quedaron asociados a este clúster que mantuvo el perfil original.

Año 2017: Áreas de Investigación ESI



Año 2018: Áreas de Investigación ESI



Año 2019: Áreas de Investigación ESI

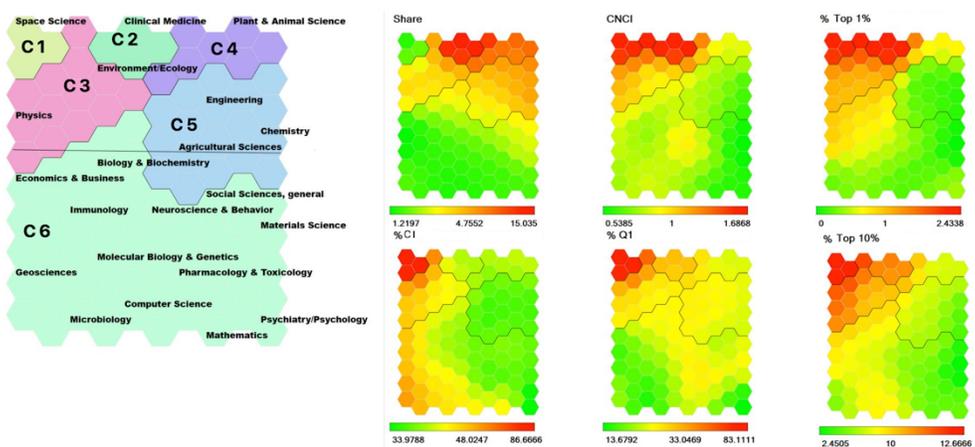


Figura 32. Mapeo neurolongitudinal de las áreas de investigación mexicanas en el *Essential Science Indicators*, 2017-2019.

### **3.7. Mapeo de la evolución del perfil cientiométrico de México.**

En esta sección presentamos la evolución del perfil cientiométrico multidimensional de México, durante 2000-2019; considerando siete indicadores: NSP, %Q1, %Top1%, %Top 10%, CNCI, %CI y CINV. En nuestra metodología para realizar este análisis introducimos una nueva técnica de mapeo, basada en redes neuronales artificiales, a la cual hemos llamado mapeo orbital. Esta técnica la realizamos con el software ViBlioSOM 2.0.

El mapa de clustering de la figura 32 muestra que, no hay cambios en la trayectoria del perfil cientiométrico multidimensional del país anualmente. Los conglomerados C1, C2, C3, C4 agrupan el perfil multidimensional de México en los primeros 10 años del período. Durante estos años, el índice de productividad nacional alcanzó los valores más altos y más 40% de los trabajos publicados se publicaron en revistas de primer cuartil. Además, 7% de los trabajos publicados en los años 2008, 2009 y 2010 se ubicaron en el 10% de lo más citado a nivel mundial. En cambio, los valores del perfil para el resto de los indicadores cientiométrico y de innovación es bajo.

En la parte superior del mapa de clustering se ubican los últimos nueve años del período de estudio. El perfil cientiométrico multidimensional de México en el año 2011, se ubica en un solo cluster (C5) por sus altos valores (rojo) en el porcentaje de trabajos de excelencia y alto desempeño, así como, valores por encima de media (naranja) en productividad, visibilidad esperada, impacto, colaboración e innovación.

El perfil del país entre 2012-2015 es muy similar, por ello estos años se localizan en el sexto cluster (C6). Durante estos años, el perfil se caracteriza por altos valores en el porcentaje de su producción de excelencia y alto desempeño, así como, valores por encima del promedio (naranja) en impacto normalizado, colaboración internacional y coeficiente de inventiva. Sin embargo, en estos tres años el perfil alcanzó valores bajos en el índice de productividad nacional y el impacto esperado.

Para finales del período (C7), el perfil alcanza los porcentajes más altos durante 2000-2019 en excelencia, impacto, colaboración internacional y coeficiente de inventiva. En cambio, el índice de productividad alcanzó valores mínimos (verde) y disminuyó el por ciento de trabajos publicados en revistas de primer cuartil.

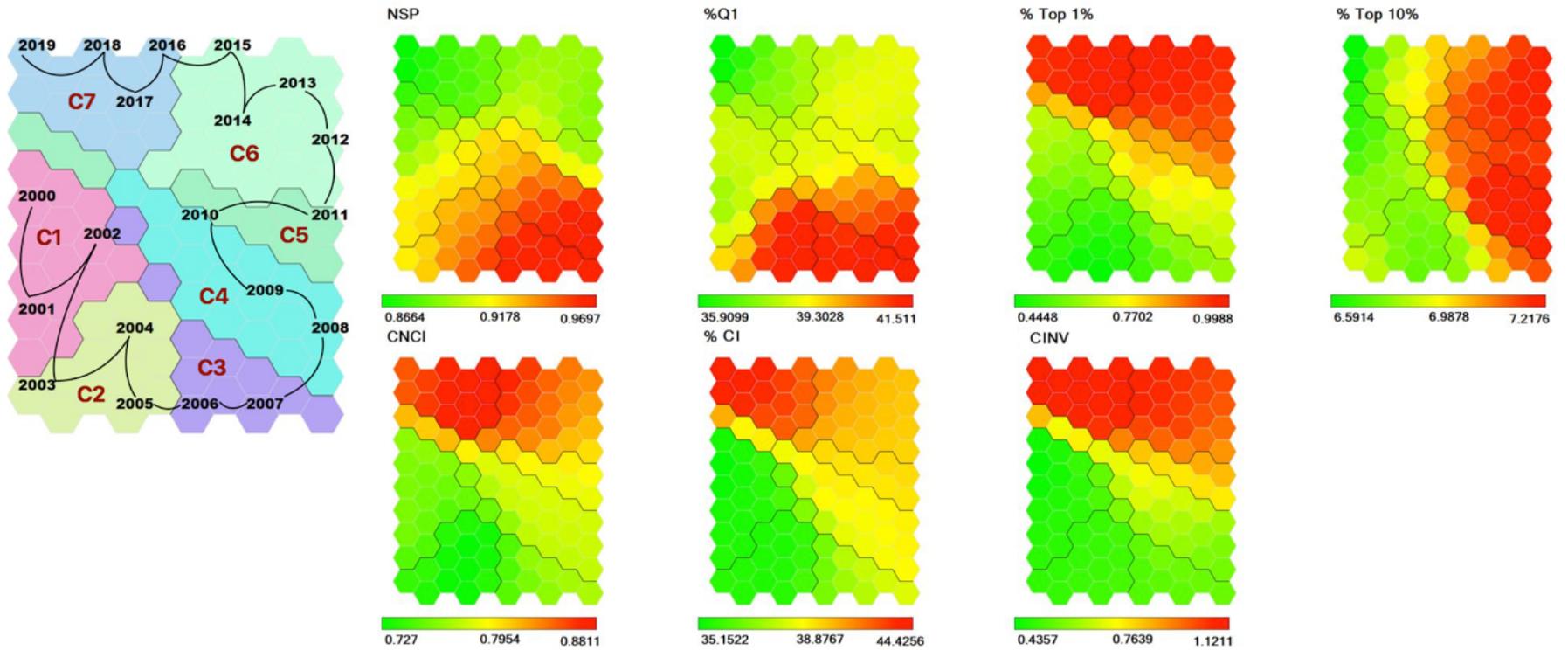


Figura 33. Mapeo de la evolución del perfil de desempeño cientométrico de México en WoS, 2000-2019.

## Capítulo IV. Consideraciones finales

Los resultados alcanzados por la presente investigación, han permitido el cumplimiento de los objetivos propuestos en cada una de sus etapas.

Por un lado, la revisión de los diversos enfoques teóricos e históricos relacionados con el objeto de la investigación, permitieron corroborar el indudable desarrollo alcanzado por las métricas de la información. Este desarrollo se ha potenciado con la perspectiva multidimensional, y la incorporación de métodos analíticos derivados de otras disciplinas y campos del conocimiento, así como nuevas técnicas de visualización. Por otro, las técnicas de frontera, y particularmente los métodos que incorporan técnicas de inteligencia artificial, han revolucionado la manera de conducir los análisis métricos, desbordando las capacidades de cómputo para procesar grandes volúmenes de datos, y construyendo modelos auto-organizados que permiten un enfoque multidimensional de los diversos problemas abordados.

En cuanto al método empleado, la propuesta documenta en forma de esquema y sobre una base metodológica todo el proceso de análisis y visualización de datos cuantitativos bidimensionales y multidimensionales. El proceso fue estructurado en cuatro etapas. La metodología presentada, la cual se ha denominado “Mapeo Neurocomputacional de la Ciencia”, es una herramienta extensible a otros profesionales y analistas de información. Con ella se puede estudiar cualquier campo de conocimiento. Sus aplicaciones pueden apoyar y complementar el estudio de sistemas complejos en el marco de los Estudios Cuantitativos de la Ciencia, la Tecnología y la Innovación, como son: sistemas de inteligencia empresarial, sistemas de vigilancia científico–tecnológicas, o sistemas de gestión de políticas científicas, entre otros.

ViBlioSOM 2.0, tecnología implementada dentro de la metodología, permitió la obtención de varios tipos de visualizaciones a partir del procesamiento de datos multidimensionales. Por tanto, el analista tiene la posibilidad de manipular e interpretar de una manera más eficiente un conjunto multidimensional de los datos. El “Mapeo Neurocomputacional de la Ciencia” coloca el concepto de perfil de desempeño multidimensional en primer plano, y proporciona recursos tecnológicos de última generación para realizar comparaciones y monitorear la evolución del perfil multidimensional de cualquier unidad de análisis.

La Metodología fue ilustrada mediante un ejemplo práctico, analizando 285,981 registros de la producción científica de México, que representan 20 años de investigaciones publicadas en el *Web of Science*.

Sin lugar a dudas, la aplicación de la metodología propuesta en contextos relacionados con ejercicios de evaluación de la investigación, tanto en México como en cualquier otro país de la región y del mundo, permitirá el análisis multidimensional

de los factores involucrados en la construcción del conocimiento científico-técnico, y su relación con el progreso científico y tecnológico alcanzado por estas naciones. En ese sentido, la metodología propuesta implica la adopción de un enfoque complejo y una visión holística que ha de tributar directamente en los procesos de toma de decisiones relacionados con las Políticas Nacionales de Ciencia, Tecnología e Innovación.

## Referencias

- Abramo, G., Andrea, C., Angelo, D., & Soldatenkova, A. (2016). The ratio of top scientists to the academic staff as an indicator of the competitive strength of universities. *Journal of Informetrics*, 10(2), 596–605. <https://doi.org/10.1016/j.joi.2016.04.013>
- Abramo, G., & D'Angelo, C. A. (2015). A methodology to compute the territorial productivity of scientists: The case of Italy. *Journal of Informetrics*, 9(4), 675–685. <https://doi.org/10.1016/j.joi.2015.07.007>
- Abramo, G., & D'Angelo, C. A. (2016). A farewell to the MNCS and like size-independent indicators. *Journal of Informetrics*, 10(2), 646–651. <https://doi.org/10.1016/j.joi.2016.04.006>
- Alcazar Farias, E., & Lozano Guzman, A. (2009). Historical development of Science and Technology Indicators advances in Latin America and Mexico. *REVISTA ESPANOLA DE DOCUMENTACION CIENTIFICA*, 32(3), 119–126. <https://doi.org/10.3989/redc.2009.3.676>
- Anderson, P. W. (2008). Theoretical paradigms for the sciences of complexity. *Lecture Notes in Physics*, 746, 229–234. [https://doi.org/10.1007/978-4-431-77056-5\\_9](https://doi.org/10.1007/978-4-431-77056-5_9)
- Arencibia-Jorge, R., Villaseñor, E. A., Lozano-Díaz, I. A., & Calvet, H. C. (2016). Elsevier's journal metrics for the identification of a mainstream journals core: A case study on Mexico. *Libres*, 26(1).
- Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11(4), 959–975. <https://doi.org/10.1016/j.joi.2017.08.007>
- Arvanitis, R., Russell, J. M., & Rosas, A. M. (1996). Experiences with the national citation reports database for measuring national performance: The case of Mexico. *Scientometrics*, 35(2), 247–255.
- Åström, F. (2006). *The Social and Intellectual Development of Library and Information Science*. Umeå University.
- Balassa, B. (1977). “Revealed” comparative advantage revisited: An analysis of relative export shares of the industrial countries, 1953–1971. *The Manchester School*, 45(4), 327–344.
- Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications*, 311(3–4), 590–614. [https://doi.org/10.1016/S0378-4371\(02\)00736-7](https://doi.org/10.1016/S0378-4371(02)00736-7)

- Barabási, A. L., & Réka, A. (1999). Emergence of scaling in random networks. *Science*, 286, 510–512. <https://doi.org/10.1126/science.286.5439.509>
- Bastian, M., Heymann, S. J., & Acomy, M. (2010). *Gephi: an open source software for exploring and manipulating networks*.
- Bath, P. A. (2004). Data mining in health and medical information. *Annual Review of Information Science and Technology*, 38, 331–369. <https://doi.org/10.1002/aris.1440380108>
- Bernal, J. D., & Mackay, A. L. (1966). Towards a Science of Science. *Zagadnienia Naukoznawstwa*, 1(2), 9.
- Björneborn, L., & Ingwersen, P. (2001). Perspectives of webometrics. *Scientometrics*, 50(1), 65–82. <https://doi.org/10.1023/A:1005642218907>
- Börner, K. (2011). Plug-and-play macroscopes. *Communications of the ACM*, 54, 60–69.
- Borner, K., & Scharnhorst, A. (2009). Visual Conceptualizations and Models of Science. Editorial for the special issue „Science of Science: Conceptualizations and Models of Science“. *Journal of Informetrics*, 3(3). Retrieved from <http://arxiv.org/abs/0903.3562>
- Borner, K, Chen, C. M., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review Of Information Science And Technology*, 37, 179–255. <https://doi.org/10.1002/aris.1440370106>
- Borner, Katy. (2010). *Atlas of science: Visualizing what we know*. MIT Press, Cambridge, MA.
- Börner, Katy, Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences of the United States of America*, 116(6), 1857–1864. <https://doi.org/10.1073/pnas.1807180116>
- Börner, Katy, Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., ... Boyack, K. W. (2012). Design and update of a classification system: The ucsd map of science. *PLoS ONE*, 7(7), 1–10. <https://doi.org/10.1371/journal.pone.0039464>
- Börner, Katy, Maru, J. T., & Goldstone, R. L. (2004). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(SUPPL. 1), 5266–5273. <https://doi.org/10.1073/pnas.0307625100>
- Börner, Katy, Scrivner, O., Cross, L. E., Gallant, M., Ma, S., Martin, A. S., ... Dilger, J. M. (2020). Mapping the co-evolution of artificial intelligence, robotics, and the

- internet of things over 20 years (1998-2017). *PLoS ONE*, 1–21. <https://doi.org/10.1371/journal.pone.0242984>
- Börner, Katy, Simpson, A. H., Bueckle, A., & Goldstone, R. L. (2018). Science map metaphors: a comparison of network versus hexmap-based visualizations. *Scientometrics*. <https://doi.org/10.1007/s11192-017-2596-3>
- Börner, Katy, Theriault, T. N., & Boyack, K. W. (2005). Mapping Science Introduction: Past , Present and Future. *Bulletin of the Association for Information Science and Technology*, 41(2), 12–16.
- Bornmann, L., Haunschild, R., & Hug, S. E. (2018). Visualizing the context of citations referencing papers published by Eugene Garfield: a new type of keyword co-occurrence analysis. *Scientometrics*, 114(2), 427–437. <https://doi.org/10.1007/s11192-017-2591-8>
- Bornmann, L., & Leydesdorff, L. (2015). Topical connections between the institutions within an organisation (institutional co-authorships, direct citation links and co-citations). *Scientometrics*, 102(1), 455–463. <https://doi.org/10.1007/s11192-014-1425-1>
- Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, 64(3), 351–374. <https://doi.org/10.1007/s11192-005-0255-6>
- Boyack, K. W., Wylie, B. N., & Davidson, G. S. (2002). Information Visualization, Human-Computer Interaction, and Cognitive Psychology: Domain Visualizations. In *Visual Interfaces to Digital Libraries*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Boyack, K. W., Wylie, B. N., & Davidson, G. S. (2015). The Global Innovation Index 2015: Effective Innovation Policies for Development, Fontainebleau, Ithaca, and Geneva. In *Stronger Innovation Linkages for*. <https://doi.org/978-2-9522210-8-5>
- Boyack, Kevin W. (2009). Using detailed maps of science to identify potential collaborations. *Scientometrics*, 79(1), 27–44. <https://doi.org/10.1007/s11192-009-0402-6>
- Boyack, Kevin W., Börner, K., & Klavans, R. (2009). Mapping the structure and evolution of chemistry research. *Scientometrics*, 79(1), 45–60. <https://doi.org/10.1007/s11192-009-0403-5>
- Boyack, Kevin W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404. <https://doi.org/10.1002/asi.21419>
- Braam, R., Moed, H., & Vanraaf, A. (1991). Mapping of Science by Combined

Cocitation and Word Analysis. II: Dynamical Aspects. *Journal of the American Society for Information Science*, 42(4), 233–251. [https://doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4<233::AID-ASI1>3.0.CO;2-I](https://doi.org/10.1002/(SICI)1097-4571(199105)42:4<233::AID-ASI1>3.0.CO;2-I)

Braun, T., Glänzel, W., & Schubert, A. (1985). Scientometric indicators. A 32 country comparison of publication productivity and citation impact. *World Scientific Publishing*, 424. Philadelphia.

Bravo Vinaja, A. (2005). Análisis bibliométrico de la producción científica de México en ciencias agrícolas a través de las bases de datos internacionales: agrícola, agris, cab abstracts, science citation index, social science citation index y tropag & rural, en el período 1983-2. Universidad Carlos III, Madrid.

Callon, M., Courtial, J. P., & 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*, 22(1), 155–205. <https://doi.org/10.1007/BF02019280>

Callon, Michel, Courtial, J. P., Turner, W. A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2), 191–235. <https://doi.org/10.1177/053901883022002003>

Cao, X., Chen, Y., & Liu, K. J. R. (2016). A data analytic approach to quantifying scientific impact. *Journal of Informetrics*, 10(2), 471–484. <https://doi.org/10.1016/j.joi.2016.02.006>

Carnap, R. (1995). *An Introduction to the Philosophy of Science*. New York: Dover Publications.

Castillo-Pérez, J. J., & Muñoz-Valera, L. (2014). Producción científica mexicana sobre influenza, 2000-2012. *Salud Pública de México*, 56(5), 424–425. Retrieved from [http://www.scielosp.org/scielo.php?script=sci\\_arttext&pid=S0036-36342014000500005&nrm=iso](http://www.scielosp.org/scielo.php?script=sci_arttext&pid=S0036-36342014000500005&nrm=iso)

Chen, C. (1999). Visualising Semantic Spaces and Author Co-Citation Networks in Digital Libraries. *Information Processing and Management*, 35, 401–420. Retrieved from [http://www.pages.drexel.edu/~cc345/papers/ip\\_m/ip\\_m.pdf](http://www.pages.drexel.edu/~cc345/papers/ip_m/ip_m.pdf)

Chen, C. (2004). Mapping Scientific Frontiers: The Quest for Knowledge Visualization. *Journal of the American Society for Information Science and Technology*, 55(4), 363–368.

Chen, C. (2006a). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of American Society for Information Science and Technology*, 57(3), 359–377. <https://doi.org/10.1002/asi>

- Chen, C. (2006b). Information Visualization: beyond the horizon. In *Journal of Chemical Information and Modeling* (Vol. 53). <https://doi.org/10.1007/1-84628-579-8>
- Chen, C. (2017). Science Mapping : A Systematic Review of the Literature. *Journal of Data and Information Science*, 2(2), 1–40. <https://doi.org/10.1515/jdis-2017-0006>
- Chen, C. (2020). *A Glimpse of the First Eight Months of the COVID-19 Literature on Microsoft Academic Graph: Themes, Citation Contexts, and Uncertainties*. Retrieved from <https://www.whitehouse.gov/briefings-statements/call-action-tech-community-new-machine-readable-covid-19->
- Chen, C., Hu, Z., Liu, S., & Tseng, H. (2012). Emerging trends in regenerative medicine: A scientometric analysis in CiteSpace. *Expert Opinion on Biological Therapy*, 12(5), 593–608. <https://doi.org/10.1517/14712598.2012.674507>
- Chen, C., Ibekwe-Sanjuan, F., & Hou, J. (2010). The Structure and Dynamics of Co-Citation Clusters: A Multiple-Perspective Co-Citation Analysis. *Journal of the American Society for Information and Technology*, 61(7), 1386–1409. Retrieved from <http://www.adsabs.harvard.edu/>
- Chen, C., & Song, M. (2017). Representing Scientific Knowledge. In *Representing Scientific Knowledge*. <https://doi.org/10.1007/978-3-319-62543-0>
- Chen, C., & Song, M. (2019). Visualizing a field of research: A methodology of systematic scientometric reviews. *PLoS ONE*, 14(10). <https://doi.org/10.1371/journal.pone.0223994>
- Chen, I.-X., & Yang, C.-Z. (2010). Visualization of Social Networks. In *Handbook of Social Network Technologies and Applications* (pp. 585–610). [https://doi.org/10.1007/978-1-4419-7142-5\\_27](https://doi.org/10.1007/978-1-4419-7142-5_27)
- Chen, Yilong, Dong, Y., Zeng, Y., Yang, X., Shen, J., Zheng, L., ... Bao, Q. (2020). Mapping of diseases from clinical medicine research—a visualization study. *Scientometrics*, 125(1), 171–185. <https://doi.org/10.1007/s11192-020-03646-8>
- Chen, Yw, Fang, S., & Börner, K. (2011). Mapping the development of scientometrics: 2002–2008. *Journal of Library Science in ...*, (2003). Retrieved from <http://ivl.slis.indiana.edu/km/pub/2011-chen-fang-borner-scienceometrics-china.pdf>
- Chinchilla-Rodriguez, Z., Zacca-Gonzalez, G., Vargas-Quesada, B., & Moya-Anegon, F. (2015). Latin American scientific output in Public Health: combined analysis using bibliometric, socioeconomic and health indicators. *Scientometrics*, 102(1), 609–628. <https://doi.org/10.1007/s11192-014-1349-9>
- Chiorni, A. I., & Guilarievkii, R. S. (2001). The impact of V . V . Nalimov on information

- science. *Scientometrics*, 52(2), 159–163.
- Chuang, Y., Lee, L., Hung, W., & Lin, P. (2010). Forging into the innovation lead a comparative analysis of scientific capacity. *International Journal of Innovation Management*, 14(3), 511–529.
- Cilliers, P. (1998). *Complexity and Postmodernism: Understanding Complex Systems*. Routledge, London, UK.
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146–166. <https://doi.org/10.1016/j.joi.2010.10.002>
- Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2012). SciMAT: A New Science Mapping Analysis Software Tool. *Journal of the Association for Information Science and Technology*, 3(68), 1609–1630.
- Cobo Martín, M. J. (2011). *SciMAT: Herramienta Software para el Análisis de la evolución del conocimiento científico. Propuesta de una Metodología de Evaluación*. Universidad de Granada.
- Collazo-Reyes, F., Luna-Morales, M. E., & Luna-Morales, E. (2017). Change in the publishing regime in Latin America: from a local to universal journal, *Archivos de investigacin Medica/Archives of Medical Research (1970-2014)*. *SCIENTOMETRICS*, 110(2), 695–709. <https://doi.org/10.1007/s11192-016-2207-8>
- Collazo-Reyes, F., Luna-Morales, M. E., Russell, J. M., & Perez-Angon, M. A. (2008). Publication and citation patterns of Latin American & Caribbean journals in the SCI and SSCI from 1995 to 2004. *Scientometrics*, 75(1), 145–161.
- Comins, J. A., & Leydesdorff, L. (2016). RPYS i/o: software demonstration of a web-based tool for the historiography and visualization of citation classics, sleeping beauties and research fronts. *Scientometrics*, 107, 1509–1517.
- Daraio, C., Fabbri, F., Gavazzi, G., Izzo, M. G., Leuzzi, L., Quaglia, G., & Ruocco, G. (2018). Assessing the interdependencies between scientific disciplinary profiles. *Scientometrics*, 116(3), 1785–1803. <https://doi.org/10.1007/s11192-018-2816-5>
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224–227.
- de Arenas, J. L., Castanos-Lomnitz, H., & Arenas-Licea, J. (2002). Significant Mexican research in the health sciences: A bibliometric analysis. *SCIENTOMETRICS*, 53(1), 39–48.

- de la Hoz-Correa, A., Muñoz-Leiva, F., & Bakucz, M. (2018). Past themes and future trends in medical tourism research: A co-word analysis. *Tourism Management*, 65, 200–211. <https://doi.org/10.1016/j.tourman.2017.10.001>
- de Solla Price, D. J. (1951). Quantitative measures of the development of science. *Archives Internationales d'Histoire Des Sciences VI e Congr`es International d'Histoire Des Sciences*, 85–93. Amsterdam.
- de Solla Price, D. J. (1963). *Little science big science*. Columbia University Press.
- del Río, J. A., Russell, J. M., & Juárez, D. (2020). Applied physics in Mexico: mining the past to predict the future. *Scientometrics*, (0123456789). <https://doi.org/10.1007/s11192-020-03639-7>
- Deliverable, W. P., Vertesy, D., & Tarantola, S. (2012). *Composite Indicators of Research Excellence*. <https://doi.org/10.2788/45492>
- Díaz-Pérez, M., Giráldez-Reyes, R., Carrillo-Calvet, H. A., Armas Peña, D., Rodríguez-Font, R. J., & Villaseñor-García, E. A. (2014). Tecnologías constituidas, innovaciones en proceso y tecnologías introducidas en el mercado internacional: caso de estudio. *TransInformação*, 23(3).
- Ding, Y., Chowdhury, G. G., & Foo, S. (2001). Bibliometric cartography of information retrieval research by using co-word analysis. *Information Processing and Management*, 37(6), 817–842. [https://doi.org/10.1016/S0306-4573\(00\)00051-0](https://doi.org/10.1016/S0306-4573(00)00051-0)
- Dwivedi, S., Garg, K. C., & Prasad, H. N. (2017). Scientometric profile of global male breast cancer research. *Current Science*, 112(9), 1814–1821. <https://doi.org/10.18520/cs/v112/i09/1814-1821>
- Egghe, L., & Rousseau, R. (1990). *Introduction to informetrics, quantitative methods in library, Documentation and information science* (Elsevier, Ed.). Amsterdam.
- Egghe, Leo, & Rousseau, R. (2002). Co-citation, bibliographic coupling and a characterization of lattice citation networks. *Scientometrics*, 55(3), 349–361. <https://doi.org/10.1023/A:1020458612014>
- Eito-Brun, R., & Ledesma Rodríguez, M. (2016). 50 years of space research in Europe: a bibliometric profile of the European Space Agency (ESA). *Scientometrics*, 109(1), 551–576. <https://doi.org/10.1007/s11192-016-2053-8>
- Evans, J., & Foster, J. (2011). Metaknowledge. *Science*, 331, 721–725.
- Fang, Y., Yin, J., & Wu, B. (2018). Climate change and tourism: a scientometric analysis using CiteSpace. *Journal of Sustainable Tourism*, 26(1), 108–126. <https://doi.org/10.1080/09669582.2017.1329310>
- Fayyad, U, Piatetsky-Shapiro, G., & Smyth, P. (1996). The KDD process for

extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11), 27–34.

Fayyad, Usama, Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in Knowledge Discovery and Data Mining*. Retrieved from citeulike-article-id:1550195

Fayyad, Usama, & Stolorz, P. (1997). Data mining and KDD: Promise and challenges. *Future Generation Computer Systems*, 13, 99–115. [https://doi.org/http://dx.doi.org/10.1016/S0167-739X\(97\)00015-0](https://doi.org/http://dx.doi.org/10.1016/S0167-739X(97)00015-0)

Fieguth, P. (2017). Concluding Thoughts. In *An Introduction to Complex Systems* (pp. 319–321). [https://doi.org/10.1007/978-3-319-44606-6\\_13](https://doi.org/10.1007/978-3-319-44606-6_13)

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., ... Barabási, A. L. (2018). Science of science. *Science*, 359(6379). <https://doi.org/10.1126/science.aao0185>

Gamarra, C., Guerrero, J. M., & Montero, E. (2016). A knowledge discovery in databases approach for industrial microgrid planning. *Renewable and Sustainable Energy Reviews*, 60, 615–630. <https://doi.org/http://dx.doi.org/10.1016/j.rser.2016.01.091>

Gantman, E. R., & Fernandez Rodriguez, C. J. (2017). Academic literature in the field of management and organizational studies in Spanish-speaking countries: A bibliometric analysis of the output in journals included in the Latindex-Catalogo database (2000-2010). *Investigacion Bibliotecologica*, 31(72), 39–61. <https://doi.org/10.22201/iibi.0187358xp.2017.57823>

García-Lillo, F., Claver, E., Marco-Lajara, B., Seva-Larrosa, P., & Ruiz-Fernández, L. (2020). MNEs from emerging markets: a review of the current literature through “bibliographic coupling” and social network analysis. *International Journal of Emerging Markets*. <https://doi.org/10.1108/IJOEM-03-2019-0170>

Garfield, E. (1979). *Citation indexing its theory and application in science, technology, and humanities*. New York: John Wiley and Sons.

Garfield, E. (1983). Mapping Science in the Third World. *Science and Public Policy*, 6(June), 112–127. <https://doi.org/10.1093/spp/10.3.112>

Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science*, 30(2), 119–145. <https://doi.org/10.1177/0165551504042802>

Garfield, E. (2009). From the science of science to Scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics*, 3(3), 173–179. <https://doi.org/10.1016/j.joi.2009.03.009>

- Garfield, E., & Sher, I. H. (1963). New factors in the evaluation of scientific literature through citation indexing. *American Documentation*, 14(3), 195–201. <https://doi.org/10.1002/asi.5090140304>
- Garfield, Eugene. (1955). Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas. *Science*, 122(3159).
- Geisler, E. (2005). The measurement of scientific activity : Research directions in linking philosophy of science. *Scientometrics*, 62(2), 269–284.
- Glänzel, W., Thijs, B., & Debackere, K. (2016, May 1). Productivity, performance, efficiency, impact-What do we measure anyway? Some comments on the paper “A farewell to the MNCS and like size-independent indicators” by Abramo and D’Angelo. *Journal of Informetrics*, Vol. 10, pp. 658–660. <https://doi.org/10.1016/j.joi.2016.04.008>
- Goldsmith, M. (1965). The science of science foundation. *Nature*, 205.
- Gordon, M. D., & Lindsay, R. K. (1996). Toward discovery support systems: A replication, re-examination, and extension of Swanson’s work on literature-based discovery of a connection between Raynaud’s and fish oil. *Journal of the American Society for Information Science*, 47(2), 116–128.
- Granovsky, Y. V. (2001). Is it possible to measure science? V . V . Nalimov ’ s research in scientometrics \*. *Scientometrics*, 52(2), 127–150.
- Grauwin, S., & Sperano, I. (2018). Bibliomaps - a software to create web-based interactive maps of science: The case of UX map. *Proceedings of the Association for Information Science and Technology*, 55(1), 815–816. <https://doi.org/10.1002/pra2.2018.14505501129>
- Guerrero, V. P., Moya-Anegón, F., & Herrero-Solana, V. (2002). Automatic extraction of relationships between terms by means of Kohonen’s algorithm. *Library and Information Science Research*, 24(3), 235–250. [https://doi.org/10.1016/S0740-8188\(02\)00124-X](https://doi.org/10.1016/S0740-8188(02)00124-X)
- Gupta, B. M., & Dhawan, S. M. (2018). Artificial intelligence research in India: A scientometric assessment of publications output during 2007-2016. *DESIDOC Journal of Library and Information Technology*, 38(6), 415–422. <https://doi.org/10.14429/djlit.38.6.12309>
- Gutiérrez-Salcedo, M., Martínez, M. Á., Moral-Munoz, J. A., Herrera-Viedma, E., & Cobo, M. J. (2018). Some bibliometric procedures for analyzing and evaluating research fields. *Applied Intelligence*, 48(5), 1275–1287. <https://doi.org/10.1007/s10489-017-1105-y>
- Guzmán Sánchez, M. V. (2009). *ViBlioSOM: Metodología para la Visualización de Información métrica con Mapas Auto-organizados*. Universidad de La Habana.

- Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804. <https://doi.org/10.1007/s11192-015-1798-9>
- He, Q. (1999). Knowledge Discovery Through Co-Word Analysis. *Library Trends*, 48(1), 133–159. <https://doi.org/x>
- Hernández-Alvarez, M., Gomez Soriano, J. M., & Martínez-Barco, P. (2017). Citation function, polarity and influence classification. *Natural Language Engineering*, 23(4), 1–28. <https://doi.org/10.1017/S1351324916000346>
- Hernández-Sampier, R., Fernández, C., & Baptista, M. (2014). *Metodología de la Investigación*.
- Herranz, N., & Ruiz-Castillo, J. (2012). Sub-field normalization in the multiplicative case: Average-based citation indicators. *Journal of Informetrics*, 6(4), 543–556. <https://doi.org/10.1016/j.joi.2012.02.006>
- Hidalgo, C. A., & Hausmann, R. (2011). The network structure of economic output. *Journal of Economic Growth*, 16(4), 309–342.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572. <https://doi.org/10.1073/pnas.0507655102>
- Hjørland, B. (2002). Domain analysis in information science: Eleven approaches - Traditional as well as innovative. *Journal of Documentation*, 58(4), 422–462. <https://doi.org/10.1108/00220410210431136>
- Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: Domain-analysis. *Journal of the American Society for Information Science*, 46(6), 400–425. [https://doi.org/10.1002/\(SICI\)1097-4571\(199507\)46:6<400::AID-ASI2>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199507)46:6<400::AID-ASI2>3.0.CO;2-Y)
- Hu, X. J., & Rousseau, R. (2009). A comparative study of the difference in research performance in biomedical fields among selected Western and Asian countries. *Scientometrics*, 81(2), 475–491. <https://doi.org/10.1007/s11192-008-2202-9>
- Huang, C., Su, J., Xie, X., Ye, X., Li, Z., Porter, A., & Li, J. (2014). A bibliometric study of China's science and technology policies: 1949–2010. *Scientometrics*, 102(2), 1521–1539. <https://doi.org/10.1007/s11192-014-1406-4>
- Hudomalj, E., & Vidmar, G. (2003). OLAP and bibliographic databases. *Scientometrics*, 58(3), 609–622. <https://doi.org/10.1023/B:SCIE.0000006883.28709.d2>
- Ibáñez, A., Larrañaga, P., & Bielza, C. (2013). Cluster methods for assessing research performance: Exploring Spanish computer science. *Scientometrics*,

97(3), 571–600. <https://doi.org/10.1007/s11192-013-0985-9>

Jaccard, P. (1901). Distribution de la flore alpine dans le Bassin des Drouces et dans quelques regions voisines. *Bulletin de La Société Vaudoise Des Sciences Naturelles*, 37(140), 241–272.

Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2 , a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLoS ONE*, 9(6), 1–12. <https://doi.org/10.1371/journal.pone.0098679>

Jimenez Andrade, J. L., Villaseñor-García, E. A., & Carrillo-Calvet, H. (2019). *LabSOM: Self Organizing Maps Laboratory*. <https://doi.org/https://doi.org/10.5281/zenodo.3630581>

Kangas, J. A., Kohonen, T. K., & Laaksonen, J. T. (1990). Variants of self-organizing maps. *IEEE Transactions on Neural Networks*, 1(1), 93–99. <https://doi.org/10.1109/72.80208>

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25.

Kim, M. C., Feng, Y., & Zhu, Y. (2020). Mapping scientific profile and knowledge diffusion of Library Hi Tech. *Library Hi Tech*. <https://doi.org/10.1108/LHT-08-2019-0164>

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, 78(9), 1464–1480. <https://doi.org/10.1109/5.58325>

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks*, 37, 52–65.

Kostoff, R N, del Rio, J. A., Cortes, H. D., Smith, C., Smith, A., Wagner, C., ... Tshiteya, R. (2007). Clustering methodologies for identifying country core competencies. *Journal of Information Science*, 33(1), 21–40. <https://doi.org/10.1177/0165551506067124>

Kostoff, Ronald N. (2002). Citation analysis of research performer quality. *Scientometrics*, 53(1), 49–71. <https://doi.org/10.1023/A:1014831920172>

Kostoff, Ronald N., & Martinez, W. L. (2005, February). Is citation normalization realistic? *Journal of Information Science*, Vol. 31, pp. 57–61. <https://doi.org/10.1177/0165551505049260>

Kuhn, T. (1996). *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Ladyman, J., Lambert, J., & Wiesner, K. (2013, January 19). What is a complex

- system? *European Journal for Philosophy of Science*, Vol. 3, pp. 33–67. <https://doi.org/10.1007/s13194-012-0056-8>
- Lancho-Barrantes, B. S., & Cantú-Ortiz, F. J. (2019). Science in Mexico: a bibliometric analysis. *Scientometrics*, (July 2018). <https://doi.org/10.1007/s11192-018-2985-2>
- Lena, M. (1997). Scientific productivity in environmental psychology in Mexico - A bibliometric analysis. *Environment and Behavior*, 29(2), 169–197.
- Leydesdorff, L. (1991). In Search of Epistemic Networks. *Social Studies of Science*, 21(1), 75–110. <https://doi.org/10.1177/030631291021001005>
- Leydesdorff, L. (1995). *The challenge of scientometrics: The development, measurement, and self-organization of scientific communications*. Leiden; Leiden University, Leiden: Science Studies/Wetenschapsstudies. DSWO Press.
- Leydesdorff, L. (1998). Theories of Citation. *Scientometrics*, 43(1), 5–25.
- Leydesdorff, L. (2008). On the normalization and visualization of author co-citation data: Salton's cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology*, 59(1), 77–85. <https://doi.org/10.1002/asi.20732>
- Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new Web-of-Science categories. *Scientometrics*, 94(2), 589–593. <https://doi.org/10.1007/s11192-012-0784-8>
- Linsay, R. K., & Gordon, M. D. (1999). Literature-based discovery by lexical statistics. *Journal of the American Society for Information Science*, 50(7), 574–587.
- Liu, B., Zhang, L., & Wang, X. (2017). Scientometric profile of global rice research during 1985-2014. *Current Science*, 112(5), 1003–1011. <https://doi.org/10.18520/cs/v112/i05/1003-1011>
- Liu, P., Chen, B. L., Liu, K., & Xie, H. (2016). Magnetic nanoparticles research: a scientometric analysis of development trends and research fronts. *Scientometrics*, 108(3), 1591–1602. <https://doi.org/10.1007/s11192-016-2017-z>
- Long, B. R. (2010). Complex science for a complex world. *Metascience*, 19(3), 441–444. <https://doi.org/10.1007/s11016-010-9427-y>
- Luis Aleixandre, J., Luis Aleixandre-Tudo, J., Bolanos-Pizarro, M., & Aleixandre-Benavent, R. (2016). Viticulture and oenology scientific research: The Old World versus the New World wine-producing countries. *INTERNATIONAL JOURNAL OF INFORMATION MANAGEMENT*, 36(3), 389–396.

<https://doi.org/10.1016/j.ijinfomgt.2016.01.003>

- Luna Morales, M. E., Luna Morales, E., & Sanchez Martinez, U. (2013). Patterns of Scientific Production and Impact of the International Maize and Wheat Improvement Center (CIMMYT) between 1966 and 2010. *INVESTIGACION BIBLIOTECOLOGICA*, 27(60), 97–124.
- Luukkonen, T., Tijssen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics*, 28(1), 15–36.
- Luwel, M. (2000). A bibliometric profile of Flemish research in natural, life and technical sciences. *Scientometrics*, 47(2), 281–302. <https://doi.org/10.1023/A:1005691010713>
- Macias-Chapula, C. A. (2013). Comparative analysis of health public policy research results among Mexico, Chile and Argentina. *SCIENTOMETRICS*, 95(2), 615–628. <https://doi.org/10.1007/s11192-012-0855-x>
- Marcum, J. W. (2001). From information center to discovery system: The next step for libraries? *Journal of Academic Librarianship*, 27(2), 97–106.
- Marshakova, I. V. (1973). A system of document connections based on references. *Scientific and Technical Information Serial of VINITI*, 6, 3–8.
- McCain, K. W. (1991). Mapping economics through the journal literature: An experiment in journal cocitation analysis. *Journal of the American Society for Information Science*, 42(4), 290–296. [https://doi.org/10.1002/\(SICI\)1097-4571\(199105\)42:4<290::AID-ASI5>3.0.CO;2-9](https://doi.org/10.1002/(SICI)1097-4571(199105)42:4<290::AID-ASI5>3.0.CO;2-9)
- Merton, R. K. (1942). *The Normative Structure of Science*.
- Mingers, J., & Leydesdorff, L. (2015). A review of theory and practice in scientometrics ☆. *European Journal of Operational Research*, 246(1), 1–19. <https://doi.org/10.1016/j.ejor.2015.04.002>
- Mistele, T., Price, T., & Hossenfelder, S. (2019). Predicting authors' citation counts and h-indices with a neural network. *Scientometrics*, 120(1), 87–104. <https://doi.org/10.1007/s11192-019-03110-2>
- Mitleton-Kelly, E. (2003). *Complex systems and evolutionary perspectives on organizations: The application of complexity theory to organizations*. Pergamon, Oxford, UK.
- Moed, H. F. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3), 265–277. <https://doi.org/10.1016/j.joi.2010.01.002>
- Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research

- impact. *Journal of the Association for Information Science and Technology*, 66(10), 1988–2002. <https://doi.org/10.1002/asi.23314>
- Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. *Profesional de La Informacion*, 29(1). <https://doi.org/10.3145/epi.2020.ene.03>
- Moral-Muñoz, J. A., López-Herrera, A. G., Herrera-Viedma, E., & Cobo, M. J. (2019). *Science mapping analysis software tools: A review*. Springer handbook of science and technology indicators.
- Moravcsik, M. J. (1980). Evaluating Science. Dimensions, Reasons and Ways. *International Conference on the Evaluation in Science and Technology. Theory and Practice Dubrovnik*.
- Moravcsik, M. J. (1985). The assesment of scientific output. *Workshop on Science and Technology Indicators in the Higher Education Sector*, 10–13. Paris.
- Moravcsik, M. J. (1989). ¿ Cómo evaluar la ciencia ya los científicos? *Revista Española de Documentación Científica*, 12(3), 313.
- Moya-Anegón, F., Herrero-Solana, V., & Jiménez-Contreras, E. (2006). A connectionist and multivariate approach to science maps: SOM, clustering and MDS applied to Library & information science research. *Journal of Information Science*, 32(1), 61–75.
- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera- Álvarez, E., Herrero-Solana, V., & Munoz-Fernández, F. (2005). Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation. *Information Processing & Management*, 41(6), 1520–1533.
- Nájera-Sánchez, J. J., Ortiz-de-Urbina-Criado, M., & Mora-Valentín, E. M. (2020). Mapping Value Co-creation Literature in the Technology and Innovation Management Field: A Bibliographic Coupling Analysis. *Frontiers in Psychology*, 11, 2376. <https://doi.org/10.3389/fpsyg.2020.588648>
- Nalimov, V. V., & Mulcsenko, E. M. (1969). *Naukometrija. Izucente rezvitja nauki kak informacionnogo processa (Scientometrics. The study of science as an information process)*. Moscú: Nauka.
- Narin, F. (1976). *Evaluative Bibliometrics: The use of publication and citation analysis in the evaluation of scientific activity*. New Jersey: Cherry Hill: Computer Horizons, Inc.
- Nicholson, S. (2002). *Bibliomining: Data Mining for Libraries*. Retrieved from <http://www.bibliomining.com/index.html>

- Nieminen, P., Pölönen, I., & Sipola, T. (2013). Research literature clustering using diffusion maps. *Journal of Informetrics*, 7(4), 874–886. <https://doi.org/10.1016/j.joi.2013.08.004>
- OECD/Eurostat. (2015). *Frascati Manual 2015*. <https://doi.org/10.1787/9789264239012-en>
- Osareh, F. (1996). Bibliometrics, citation analysis and co-Citation Analysis: A review of literature I. *Libri*, 46, 149–158.
- Özmutlu, S., & Çavdur, F. (2005). Neural network applications for automatic new topic identification. *Online Information Review*, 29(1), 34–53. <https://doi.org/10.1108/14684520510583936>
- Palvia, P., Serenko, A., Benitez, J., & Liang, T.-P. (2020). Professor Edgar H. Sibley: A Tribute and A Scientometric Profile. *Journal of Global Information Technology*, 23(4), 241–247. <https://doi.org/10.1080/1097198X.2020.1834069>
- Pan, X., Yan, E., Cui, M., & Hua, W. (2018). Examining the usage, citation, and diffusion patterns of bibliometric mapping software: A comparative study of three tools. *Journal of Informetrics*, 12(2), 481–493. <https://doi.org/10.1016/j.joi.2018.03.005>
- Payette, N. (2012). Agent-based models of science. In A. Scharnhorst, K. Börner, & P. van den Besselaar (Eds.), *Models of science dynamics* (pp. 127–158). Berlin: Springer.
- Pereira, J. C. R., Escuder, M. M. L., & Zanetta, D. M. T. (1998). Brazilian sciences and government funding at the State of São Paulo. *Scientometrics*, 43(2), 177–188. <https://doi.org/10.1007/BF02458406>
- Perez-Angon, A. M. (2018). *Atlas de la Ciencia Mexicana*. Retrieved from <http://www.atlasdelacienciamexicana.org/es/indicadores.html>
- Perez-Santos, J. L. M., & Anaya-Ruiz, M. (2013). Mexican Breast Cancer Research Output, 2003-2012. *Asian Pacific Journal of Cancer Prevention*, 14(10), 5921–5923. <https://doi.org/10.7314/apjcp.2013.14.10.5921>
- Persson, O., R., Danell, J., & Wiborg, S. (2009). *How to use Bibexcel for various types of bibliometric analysis*.
- Persson, O. (1994). The intellectual base and research fronts of JASIS 1986–1990. *Journal of the American Society for Information Science*, 45(1), 31–38. [https://doi.org/10.1002/\(SICI\)1097-4571\(199401\)45:1<31::AID-ASI4>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-4571(199401)45:1<31::AID-ASI4>3.0.CO;2-G)
- Peters, H. P. F., & van Raan, A. F. J. (1993a). Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional

- scaling. *Research Policy*, 22(1), 23–45. [https://doi.org/10.1016/0048-7333\(93\)90031-C](https://doi.org/10.1016/0048-7333(93)90031-C)
- Peters, H. P. F., & van Raan, A. F. J. (1993b). Co-word-based science maps of chemical engineering. Part II: Representations by combined clustering and multidimensional scaling. *Research Policy*, 22(1), 47–71. [https://doi.org/10.1016/0048-7333\(93\)90032-D](https://doi.org/10.1016/0048-7333(93)90032-D)
- Peters, H. P. F., & van Raan, A. F. J. (1994). A bibliometric profile of top-scientists a case study in chemical engineering. *Scientometrics*, 29(1), 115–136. <https://doi.org/10.1007/BF02018386>
- Pritchard, A. (1969). Statistical Bibliography or Bibliometrics. *Journal of Documentation*, 25(4).
- R Core Team. (2016). *R: A language and environment for statistical computing*. Retrieved from <https://www.r-project.org>
- Ravikumar, S., Agrahari, A., & Singh, S. N. (2015). Mapping the intellectual structure of scientometrics: A co-word analysis of the journal scientometrics (2005–2010). *Scientometrics*, 102(1), 929–955. <https://doi.org/10.1007/s11192-014-1402-8>
- RICYT. (2019). Los Investigadores Universitarios y su Vínculo con el Entorno en América Latina. *El Estado de La Ciencia*, 30(2), 29–40.
- Russell, J. M., Hernandez-Garcia, Y., & Kleiche-Dray, M. (2016). Collaboration dynamics of Mexican research in Chemistry and its relationship with communication patterns. *SCIENTOMETRICS*, 109(1), 283–316. <https://doi.org/10.1007/s11192-016-2069-0>
- Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Sancho, R. (1990). Indicadores bibliométricos utilizados en la evaluación de la ciencia y la tecnología. Revisión bibliográfica —. *Revista Española de Documentación Científica*, 13(3–4), 842–865.
- Scharnhorst, A., Borner, K., & van den Besselaar, P. (2012). *Model of Science Dynamics* (Springer, Ed.). New York.
- Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9(5–6), 281–291. <https://doi.org/10.1007/BF02017249>
- Schulz, P. A., & Manganote, E. J. T. (2012). Revisiting country research profiles: Learning about the scientific cultures. *Scientometrics*, 93(2), 517–531. <https://doi.org/10.1007/s11192-012-0696-7>

- Schwab, K. (2019). *The Global Competitiveness Report 2019*. Retrieved from [http://www3.weforum.org/docs/WEF\\_TheGlobalCompetitivenessReport2019.pdf](http://www3.weforum.org/docs/WEF_TheGlobalCompetitivenessReport2019.pdf)
- SCIMAGO. (2018). *Principales indicadores cientométricos de la actividad investigadora*.
- ScimagoLab. (2007). *Scimago Journal & Country Rank*.
- Selin, H. (2006). *Encyclopedia of the History of Science, Technology, and Medicine in Non-Western Cultures*. Berlin: Springer.
- Shao, Y., Hu, Z., Luo, M., Huo, T., & Zhao, Q. (2020). What is the policy focus for tourism recovery after the outbreak of COVID-19? A co-word analysis. *Current Issues in Tourism*. <https://doi.org/10.1080/13683500.2020.1806798>
- Sheehan, H. M. (2007). J D Bernal: Philosophy, politics and the science of science. *Journal of Physics: Conference Series*, 57(1), 29–39. <https://doi.org/10.1088/1742-6596/57/1/003>
- Sideris, G., Katsaros, D., Sidiropoulos, A., & Manolopoulos, Y. (2018). The Science of Science and a Multilayer Network Approach to Scientists ' Ranking. *IDEAS 2018: 22nd International Database Engineering & Applications Symposium*, 7. Villa San Giovanni, Italy: ACM, New York, NY, USA.
- Sierra-Flores, M. M., Guzmán, M. V., Raga, A. C., & Pérez, I. (2009). The productivity of Mexican astronomers in the field of outflows from young stars. *Scientometrics*, 83(1), 765–777.
- Skalska-Zlat, M. (2001). Nalimov and the Polish way towards science of science. *Scientometrics*, 52(2), 211–223. <https://doi.org/10.1023/A:1017911722525>
- Skupin, A., Biberstine, J. R., & Börner, K. (2013). Visualizing the Topical Structure of the Medical Sciences: A Self-Organizing Map Approach. *PLoS ONE*, 8(3). <https://doi.org/10.1371/journal.pone.0058779>
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269. <https://doi.org/10.1002/asi.4630240406>
- Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, 38(2), 275–293. <https://doi.org/10.1007/BF02457414>
- Small, H. (1998). Citations and consilience in science. *Scientometrics*, 43(1), 143–148. <https://doi.org/10.1007/bf02458403>
- Small, Henry. (1981). The relationship of information science to the social sciences: A co-citation analysis. *Information Processing and Management*, 17(1), 39–50.

[https://doi.org/10.1016/0306-4573\(81\)90040-6](https://doi.org/10.1016/0306-4573(81)90040-6)

Small, Henry, Boyack, K. W., & Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, 43(8), 1450–1467. <https://doi.org/10.1016/j.respol.2014.02.005>

Soós, S. (2011). The functional anatomy of science mapping: Katy Börner: Atlas of science: Visualizing what we know. *Scientometrics*, 89(2), 723–726. <https://doi.org/10.1007/s11192-011-0480-0>

Sotolongo Aguilar, G., Guzmán Sánchez, M. M. V., & Carrillo, H. (2002). ViblioSom: Visualización de información bibliométrica mediante el mapeo autoorganizado. *Revista Española de Documentación Científica*, 25(4), 477–484. Retrieved from <http://redc.revistas.csic.es/index.php/redc/article/view/281/337>  
<http://80stas.csic.es/index.php/redc/article/view/281/337>

Šubelj, L., Van Eck, N. J., & Waltman, L. (2016). Clustering scientific publications based on citation relations: A systematic comparison of different methods. *PLoS ONE*, 11(4). <https://doi.org/10.1371/journal.pone.0154404>

Swanson, D. R. (1986). Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine*, 30(1), 7–18.

Thelwall, M., & Levitt, J. M. (2018). National scientific performance evolution patterns: Retrenchment, successful expansion, or overextension. *Journal of the Association for Information Science and Technology*, 69(5), 720–727. <https://doi.org/10.1002/asi.23969>

Thijs, B. (2020). Using neural-network based paragraph embeddings for the calculation of within and between document similarities. *Scientometrics*, 125, 835–849. <https://doi.org/10.1007/s11192-020-03583-6>

Thijs, B., & Glänzela, W. (2009). A structural analysis of benchmarks on different bibliometrical indicators for European research institutes based on their research profile. *Scientometrics*, 79(2), 377–388. <https://doi.org/10.1007/s11192-009-0425-z>

Thor, A., Marx, W., Leydesdorff, L., & Bornmann, L. (2016). Introducing CitedReferencesExplorer ( CRExplorer ): A program for reference publication year spectroscopy with cited references standardization. *Journal of Informetrics*, 10(2), 503–515. <https://doi.org/10.1016/j.joi.2016.02.005>

Turnpenny, J., Jones, M., & Lorenzoni, I. (2011). *Where Now for Science ? : A Critical Review of its Development , Definitions , and Uses*. 36(3), 287–306. <https://doi.org/10.1177/0162243910385789>

Van Den Besselaar, P., & Leydesdorff, L. (1996). Mapping change in scientific specialties: A scientometric reconstruction of the development of artificial

- intelligence. *Journal of the American Society for Information Science*, 47(6), 415–436. [https://doi.org/10.1002/\(SICI\)1097-4571\(199606\)47:6<415::AID-ASIS3>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-4571(199606)47:6<415::AID-ASIS3>3.0.CO;2-Y)
- Van Eck, N. J., & Waltman, L. (2014). CitNetExplorer: A new software tool for analyzing and visualizing citation networks. *Journal of Informetrics*, 8, 802–823.
- van Eck, Nees Jan. (2011). *Methodological advances in bibliometric mapping of science* (Erasmus University Rotterdam). Retrieved from <http://repub.eur.nl/res/pub/26509/>
- van Eck, Nees Jan, & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538. <https://doi.org/10.1007/s11192-009-0146-3>
- van Eck, Nees Jan, & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, 111(2), 1053–1070. <https://doi.org/10.1007/s11192-017-2300-7>
- Van Eck, Néés Jan, & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635–1651. <https://doi.org/10.1002/asi.21075>
- Van Eck, Néés Jan, & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring Scholarly Impact*. <https://doi.org/10.1007/978-3-319-10377-8>
- Vera-Baceta, M. A., Thelwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics*, 121, 1803–1813. <https://doi.org/10.1007/s11192-019-03264-z>
- Vilchez-Román, C., & Quiliano-Terreros, R. (2018). Stability and change in public health studies in Colombia and Mexico: An exploratory approach based on co-word analysis. *Revista Panamericana de Salud Publica/Pan American Journal of Public Health*, Vol. 42. <https://doi.org/10.26633/rpsp.2018.35>
- Villaseñor, E. A., Carrillo-Calvet, H., & Arencibia-Jorge, R. (2017). Multiparametric characterization of scientometric performance profiles assisted by neural networks: a study of Mexican higher education institutions. *Scientometrics*, 109(2). <https://doi.org/10.1007/s11192-016-2166-0>
- Waltman, L. (2016). Special section on size-independent indicators in citation analysis. *Journal of Informetrics*, 10(2), 645. <https://doi.org/10.1016/j.joi.2016.04.001>
- Waltman, L., & van Eck, N. J. (2007). Some Comments on the Question Whether Co-Occurrence Data Should Be Normalized. *Journal of American Society for*

*Information Science and Technology*, 58(11), 1701–1703.

Waltman, L., & van Eck, N. J. (2013). Source normalized indicators of citation impact: An overview of different approaches and an empirical comparison. *Scientometrics*, 96(3), 699–716. <https://doi.org/10.1007/s11192-012-0913-4>

Waltman, L., & van Eck, N. J. (2015). Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, 9(4), 872–894. <https://doi.org/10.1016/j.joi.2015.08.001>

Waltman, L., van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635. <https://doi.org/10.1016/j.joi.2010.07.002>

Wei, F., Grubestic, T. H., & Wade Bishop, B. (2015). Exploring the GIS Knowledge Domain Using CiteSpace. *The Professional Geographer*, 67(3), 374–384. <https://doi.org/10.1080/00330124.2014.983588>

White, H. D. (1990). Author co-citation analysis: Overview and defense. In *Scholarly communication and bibliometrics* (C. L. Borg, pp. 84–106). Newbury Park, CA.

White, Howard D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163–171. <https://doi.org/10.1002/asi.4630320302>

White, Howard D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science*, 49(4), 327–355. [https://doi.org/10.1002/\(SICI\)1097-4571\(19980401\)49:4<327::AID-ASI4>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1097-4571(19980401)49:4<327::AID-ASI4>3.0.CO;2-W)

Willett, & Restrepo. (2017). The Journal of Mathematical Chemistry: A Bibliometric Profile. *Journal of Mathematical Chemistry*, 55, 1589–1596. <https://doi.org/10.1007/s10910-017-0747-7>

World Bank Group. (2021). *World Bank Open Data*.

Wouters, P. (1999). *The Citation Culture*. University of Amsterdam.

Wu, Y., Wang, H., Wang, Z., Zhang, B., & Meyer, B. C. (2020). Knowledge mapping analysis of rural landscape using CiteSpace. *Sustainability (Switzerland)*, 12(1), 1–17. <https://doi.org/10.3390/SU12010066>

Yuexiao, Z. (1988). Definitions and science of information. *Information Processing & Management*, 24(4), 479–491.

Zhang, M. (2010). Social Network Analysis: History, Concepts, and Research. In *Handbook of Social Network Technologies and Applications* (pp. 3–21).

[https://doi.org/10.1007/978-1-4419-7142-5\\_1](https://doi.org/10.1007/978-1-4419-7142-5_1)

- Zhao, D., & Strotmann, A. (2008). Information science during the first decade of the web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology*, 59(6), 916–937. <https://doi.org/10.1002/asi.20799>
- Zhou, P., & Leydesdorff, L. (2010). *Fractional counting of citations in research evaluation: An option for cross- and interdisciplinary assessments*. 1–26. Retrieved from <http://arxiv.org/abs/1012.0359>
- Ziman, J. (2000). *Real Science: What it Is and What it Means*. Cambridge University Press.
- Zitt, M. (2010). Citing-side normalization of journal impact: A robust variant of the Audience Factor. *Journal of Informetrics*, 4(3), 392–406. <https://doi.org/10.1016/j.joi.2010.03.004>

## Anexos

### Anexo 1: Lista de acrónimos de las 60 instituciones mexicanas más productivas en WoS

Institución	Sigla
Benemerita Universidad Autonoma de Puebla	BUAP
Centro de Investigacion & Desarrollo Tecnologico en Electroquimica S.C.	CIDETEQ
Centro de Investigacion Cientifica de Yucatan	CICY
Centro de Investigacion Cientifica y de Educacion Superior de Ensenada	CICESE
Centro de Investigacion en Alimentacion y Desarrollo	CIAD
Centro de Investigacion en Matematicas	CIMAT
Centro de Investigacion en Materiales Avanzados S.C.	CIMAV
Centro de Investigacion Quimica Aplicada	CIQA
Centro de Investigacion y de Estudios Avanzados del Instituto Politecnico Nacional	CINVESTAV
Centro de Investigacion y Docencia Económicas	CIDE
Centro de Investigaciones en Optica A.C.	CIO
Centro Internacional de Mejoramiento del Maiz y el Trigo	CIMMYT
Colegio de Mexico	COLMEX
Colegio de Postgraduados - Mexico	COLPOS
El Colegio de la Frontera Sur	ECOSUR
Hospital General de Mexico	HGM
Hospital Infantil de Mexico Federico Gomez	HIMFG
Instituto de Cardiología	INC
Instituto de Ecología - Mexico	INECC
Instituto Mexicano del Petroleo	IMP
Instituto Mexicano del Seguro Social	IMSS
Instituto Nacional de Astrofisica, Optica y Electronica INAOE	INAOE
Instituto Nacional de Cancerologia	INCAN
Instituto Nacional de Ciencias Medicas y Nutricion Salvador Zubiran - Mexico	INCMN
Instituto Nacional de Medicina Genomica	INMEGEN
Instituto Nacional de Psiquiatria Ramon de la Fuente Muniz	INPRFR
Instituto Nacional de Salud Publica	INSP
Instituto Politecnico Nacional - Mexico	IPN
Instituto Potosino Investigacion Cientifica y Tecnologica	IPICT
Instituto Tecnologico Autonomo de Mexico	ITAM
Tecnologico de Monterrey	ITESM
Universidad Americas Puebla	UDLAP
Universidad Anahuac	UA
Universidad Autonoma de Aguascalientes	UAA
Universidad Autonoma de Baja California	UABC
Universidad Autonoma de Chiapas	UNACH

Universidad Autonoma de Ciudad Juarez	UACJ
Universidad Autonoma de Coahuila	UAdeC
Universidad Autonoma de Nuevo Leon	UANL
Universidad Autonoma de Queretaro	UAQ
Universidad Autonoma de San Luis Potosi	UASP
Universidad Autonoma de Sinaloa UAS	UAS
Universidad Autonoma de Tamaulipas	UAT
Universidad Autonoma de Yucatan	UADY
Universidad Autonoma de Zacatecas	UAZ
Universidad Autonoma del Estado de Hidalgo	UAEH
Universidad Autonoma del Estado de Mexico	UAEM
Universidad Autonoma del Estado de Morelos	UAEM
Universidad Autonoma Metropolitana - Mexico	UAM
Universidad de Colima	U de C
Universidad de Guadalajara	UdeG
Universidad de Guanajuato	UG
Universidad de Monterrey	UDEM
Universidad de Sonora	UNISON
Universidad Juarez Autonoma de Tabasco	UJAT
Universidad Juarez del Estado de Durango	UJED
Universidad Michoacana de San Nicolas de Hidalgo	UMSNH
Universidad Nacional Autonoma de Mexico	UNAM
Universidad Panamericana - Ciudad de Mexico	UP
Universidad Veracruzana	UV

Anexo 2: Lista de acrónimos de las instituciones extranjeras con mayor número de publicaciones en colaboración con México.

<b>Organizaciones</b>	<b>Siglas</b>
University of California System	UC
University of Texas System	UT
State University System of Florida	SUSF
United States Department of Energy	DOE
Harvard University	HU
University of California San Diego	UCSan Diego
Texas A&M University System	Texas A&M
State University of New York System	SUNY
University of Illinois System	UIC
Texas A&M University College Station	Texas A&M
University of California Davis	ucd
University of Chicago	UChicago
University of California Los Angeles	UCLA
University of Wisconsin System	UW
Ohio State University	OSU
University of Florida	UF
University System of Maryland	USM
Johns Hopkins University	JHU
University of Wisconsin Madison	UWM
University of Arizona	UA
University of Colorado System	CU
University of California Riverside	UCR
Northwestern University	UN
Cornell University	CU
Purdue University System	PUS
Centre National de la Recherche Scientifique	CNRS
Universite Paris Saclay	UPS
CEA	CEA
National Institute of Nuclear and Particle Physics	IN2P3
Sorbonne Universite	SU
Universites de Strasbourg Etablissements Associes	USEA
Universite Claude Bernard Lyon 1	UCBL1
Russian Academy of Sciences	RAS
Lomonosov Moscow State University	LMSU
National Research Centre - Kurchatov Institute	NRCKI
Joint Institute for Nuclear Research - Russia	JINR

Alikhanov Institute for Theoretical & Experimental Physics	AITEP
Consejo Superior de Investigaciones Cientificas	CSIC
University of Barcelona	UB
Autonomous University of Madrid	UAM
Istituto Nazionale di Fisica Nucleare	INFN
Sapienza University Rome	SU
University of Padua	UP
Helmholtz Association	HA
Max Planck Society	MPG
Imperial College London	ICL
University of London	UCL
Chinese Academy of Sciences	CAS
Universidade de Sao Paulo	USP
Charles University Prague	CUNI

Anexo 3: Producción en WoS, Población total en 2019, Producción per cápita, *Category Normalized Citation Impact*, % %Documentos en el Top 1%, %Documentos en el Top 10%, % Colaboración internacional, % Documentos publicados en Q1y el % de Autores para Correspondencia de los 10 países Latinoamericanos más productivos en WoS.

Países	Ndoc WoS	Pob 2019	Prod x 10000 Hab	CNCI	% Top 1%	% Top 10%	% CI	% Q1	% A Corr
BRAZIL	205866	211049527	9.754392863	0.89113927	0.87386941	7.29163631	37.7313398	37.53011116	76.1179602
MEXICO	69531	127575529	5.450183162	0.84188432	0.94490227	6.58123715	44.4348564	35.84534069	69.5646546
ARGENTINA	40868	44938712	9.094163624	0.97752391	1.25770774	7.57805618	48.1305667	43.41068181	63.9987276
CHILE	39841	18952038	21.02201357	1.20641356	1.54363595	10.3913054	64.3834241	47.9852201	57.7018649
COLOMBIA	28755	50339443	5.712220535	0.98344745	1.52669101	8.25595549	56.6718832	43.13077029	58.3168145
PERU	8121	32510453	2.497965808	1.24433461	2.01945573	9.87563108	71.3705209	42.41771763	41.1525674
ECUADOR	9089	17373662	5.231481998	1.08700589	1.70535813	8.86786225	72.9013093	43.07768106	51.5458246
URUGUAY	5163	3461734	14.91449083	1.07149206	1.25895797	8.11543676	69.2620569	42.87932564	50.2033701
VENEZUELA	3447	28515829	1.208802311	1.02419997	1.76965477	7.80388744	75.5149405	35.72156197	45.1987235
CUBA	3408	11333483	3.007019113	0.77080214	0.85093897	6.19131455	78.9319249	33.13650573	43.7793427

Anexo 4: México y 29 países con similar desempeño socioeconómico: Población total en 2019, Producto Interno Bruto en 2019, *Global Competitiveness Index 2019, Economic Complexity Index 2018*.

Países	Siglas	Pob 2019	PIB 2019	ICG 2019	ECI 2018
Finland	FIN	5520314	269296310180.39	96.3866154	12
Bahrain	BHR	1641172	38574069148.94	92.7825213	48
Singapore	SGP	5703569	372062527488.64	92.5877969	5
Estonia	EST	1326590	31471100656.24	89.7537232	28
Switzerland	CHE	8574832	703082435158.86	84.7970804	2
Portugal	PRT	10269417	238785085030.99	84.5009009	33
Norway	NOR	5347896	403336363636.36	83.405749	43
Lithuania	LTU	2786844	54627411860.47	82.3604345	31
Japan	JPN	126264931	5081769542379.77	82.1381966	1
Spain	ESP	47076781	1393490524517.64	81.2083006	32
Saudi Arabia	SAU	34268528	792966838161.66	80.991753	36
Austria	AUT	8877067	445075391688.16	79.5675675	6
Belgium	BEL	11484055	533097455834.51	78.3794324	21
Netherlands	NLD	17332850	907050863145.10	78.3328136	27
Slovenia	SVN	2087946	54174227308.89	76.0825714	10
Ireland	IRL	4941444	388698711348.16	73.2191404	17
Denmark	DNK	5818553	350104327658.68	71.2371985	24
Latvia	LVA	1912789	34102913582.42	70.7105001	34
Canada	CAN	37589262	1736425629519.96	70.5308199	39
Sweden	SWE	10285453	530883869004.98	70.3529199	8
Israel	ISR	9053300	394652210295.85	70.3502814	20
Poland	POL	37970874	595858207011.51	68.5731173	23
Mexico	MEX	127575529	1268870527160.03	64.9	19
Romania	ROU	19356544	250077444017.08	63.4432157	25
United Kingdom	GBR	66834405	2829108219165.80	61.7064238	13
Germany	DEU	83132799	3861123558039.21	60.1168394	4
Hungary	HUN	9769949	163469042867.96	58.7712129	9
Thailand	THA	69625582	543548969968.69	54.3111165	22
Italy	ITA	60297396	2003576145498.04	45.0685024	14

Anexo 5. Producción en WoS, Población total en 2019, Producción per cápita, *Category Normalized Citation Impact*, % Documentos en el Top 1%, % Documentos en el Top 10%, % Colaboración internacional, % Documentos publicados en Q1y el % de Autores para Correspondencia para México y 28 países.

Países	Ndoc WoS	Pob 2019	Prod x 10000	CNCI	% Top 1%	% Top 10%	% CI	% Q1	% A Corr
AUSTRIA	82999	8877067	93.49822413	1.418373	2.21087001	13.7399246	65.4104266	53.6099428	51.504235
BAHRAIN	1324	1641172	8.067405488	1.952596	2.64350453	8.4592145	68.2779456	31.6067653	48.4138973
BELGIUM	110395	11484055	96.12893703	1.60135626	2.42492867	15.1528602	67.3717107	56.2948928	50.7649803
CANADA	347544	37589262	92.45831961	1.36901131	1.93040306	12.7566581	52.7691458	55.2956669	60.3080473
DENMARK	92808	5818553	159.5035742	1.59535801	2.45991725	15.8951814	64.2509266	59.5777805	52.8930696
ESTONIA	9768	1326590	73.63239584	1.71750835	3.11220311	15.4074529	66.8304668	50.9284177	52.54914
FINLAND	63024	5520314	114.1674187	1.40812039	2.04207921	14.0010155	62.1128459	54.9978432	57.7240416
GERMANY	534826	83132799	64.33393395	1.30813595	1.73383493	12.6287802	52.8532644	52.6850989	62.6650537
HUNGARY	34505	9769949	35.31748221	1.31544257	1.97942327	10.9375453	55.4354441	45.6732608	57.3627011
IRELAND	53013	4941444	107.2824057	1.33870181	2.01648652	12.2422802	54.5526569	54.2578742	48.478675
ISRAEL	69489	9053300	76.75543724	1.33291084	1.88231231	12.3976457	49.8510556	54.6371037	63.3668638
ITALY	349682	60297396	57.9928858	1.33094368	1.71470079	13.0887492	49.238165	51.6514542	64.8946757
JAPAN	377275	126264931	29.87963459	0.9226687	0.97965675	7.58120734	30.4965874	44.039474	72.22583
LATVIA	7059	1912789	36.90422728	1.36899159	2.13911319	11.4605468	46.6213345	48.5261402	61.4109647
LITHUANIA	11946	2786844	42.86569324	1.04968516	1.57374854	9.68525029	51.1468274	40.0384164	61.9705341
MEXICO	69772	127575529	5.469073932	0.84422116	0.95167116	6.61440119	44.4203979	35.8665119	69.5135584
NETHERLANDS	198443	17332850	114.4895387	1.60163746	2.54481136	15.9552113	62.1634424	60.5541272	54.1026894
NORWAY	69143	5347896	129.2900984	1.48425979	2.34152409	14.4786891	62.5500774	53.6868385	56.5364534
POLAND	137638	37970874	36.24830969	1.02314798	1.13413447	8.36324271	35.3645069	37.4300151	74.9720281
PORTUGAL	78763	10269417	76.69666155	1.18550048	1.53371507	11.1067379	53.0617168	52.3922524	60.3100441
ROMANIA	49902	19356544	25.78042857	0.94919846	1.22640375	7.73115306	33.5417418	32.6109255	73.4980562
SAUDI ARABIA	61718	34268528	18.0101112	1.34686623	2.66534884	14.9372954	76.4914612	40.3555067	52.2278752

SINGAPORE	67834	5703569	118.9325491	1.63200297	2.83780995	17.3673969	65.0676652	61.0511757	58.1581508
SLOVENIA	17884	2087946	85.65355617	1.21913943	1.73339298	11.2726459	56.3632297	45.933205	55.7481548
SPAIN	288766	47076781	61.33936813	1.26467221	1.56701274	11.5616104	50.7816017	55.3011347	62.9544337
SWEDEN	132133	10285453	128.4659023	1.46242068	2.23713985	14.7835893	65.0829089	56.9711282	53.9184004
SWITZERLAND	155612	8574832	181.4752756	1.65863057	2.67074519	16.4781636	69.9065625	58.7895368	48.5316042
THAILAND	42826	69625582	6.15090011	0.91850275	1.13015458	7.97879793	46.8874048	39.3113213	67.6831831
UNITED KINGDOM	671669	66834405	100.4974908	1.42841703	1.99383923	13.6224837	53.3282018	58.2919387	58.3348346

Anexo 6. Indicadores cuantitativos que caracterizan el perfil de desempeño de las instituciones mexicanas más productivas en el período 2015-2019 en WoS.

Ndoc WoS: Producción científica mexicana

ISP: Institutional Scientific Productivity

CNCI: *Category Normalized Citation Impact*

%Q1: Porcentaje de documentos publicados en revistas de cuartil 1

%Top 1%: Porcentaje de la producción científica mexicana en el 1% de lo más citado

% Top 10%: Porcentaje de la producción científica mexicana en el 10% de lo más citado

% CI: Porcentaje de documentos en colaboración Internacional

% Acorr: Porcentaje de la producción con autores mexicanos como autor para correspondencia

Instituciones	Siglas	Ndoc WoS	ISP	CNCI	% Q1	% Top 1%	% Top 10%	% CI	% ACorr
Universidad Nacional Autónoma de México	UNAM	25512	5.61	0.84	37.48	0.80	6.11	45.53	50.35
Instituto Politécnico Nacional - México	IPN	10184	8.55	0.78	32.36	0.70	5.50	35.21	50.88
Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional	CINVESTAV	6188	8.48	1.14	43.33	1.33	9.11	45.67	41.89
Universidad Autónoma Metropolitana - México	UAM	4313	3.87	1.12	28.45	0.81	4.43	30.33	48.41
Universidad Autónoma de Nuevo León	UANL	3829	4.75	0.75	33.57	0.60	5.22	38.29	47.69
Universidad de Guadalajara	UdeG	3663	3.09	0.68	29.30	0.57	4.07	37.59	42.72
Tecnológico de Monterrey	ITESM	3436	6.76	1.11	42.58	1.16	10.30	53.17	46.89
Benemérita Universidad Autónoma de Puebla	BUAP	3292	4.96	1.23	40.47	2.04	11.76	51.61	36.60
Universidad Autónoma de San Luis Potosí	UASP	2937	5.93	1.07	41.12	1.77	10.15	47.46	39.80
Instituto Mexicano del Seguro Social	IMSS	2645	8.29	0.88	30.01	0.83	5.10	29.34	31.23
Universidad de Guanajuato	UG	2207	4.16	0.76	37.73	0.41	6.07	42.86	51.11
Universidad Autónoma del Estado de Morelos	UAEM	2049	6.13	0.57	28.10	0.15	3.12	34.21	47.68
Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán - México	INCMN	1944	10.92	1.70	45.96	1.95	9.47	38.07	33.64

Universidad Veracruzana	UV	1909	4.19	0.62	25.25	0.26	3.30	35.46	43.79
Universidad Michoacana de San Nicolas de Hidalgo	UMSNH	1863	5.28	0.82	34.75	0.75	5.85	43.53	48.15
Instituto Nacional de Salud Publica	INSP	1845	10.08	3.33	48.45	4.23	13.82	63.36	34.80
Universidad Autonoma de Baja California	UABC	1773	4.09	0.64	26.76	0.45	4.46	42.58	41.46
Universidad Autonoma del Estado de Mexico	UAEM	1658	3.24	0.60	20.93	0.00	4.52	34.56	53.32
Instituto de Ecologia - Mexico	INECC	1512	11.91	0.81	27.95	0.66	6.15	47.29	48.35
Universidad de Sonora	UNISON	1447	4.32	0.69	32.35	0.28	4.98	39.46	46.44
Colegio de Postgraduados - Mexico	COLPOS	1406	5.41	0.48	16.06	0.28	3.41	24.96	52.56
Centro de Investigacion Cientifica y de Educacion Superior de Ensenada	CICESE	1355	6.71	0.73	36.81	0.52	4.65	45.17	48.63
Universidad Autonoma de Queretaro	UAQ	1240	3.95	0.73	33.77	0.24	5.40	33.06	48.95
Universidad Autonoma del Estado de Hidalgo	UAEH	1226	3.82	0.69	25.37	0.57	3.43	29.61	53.10
Universidad Autonoma de Yucatan	UADY	1202	4.36	0.66	28.22	0.33	4.74	43.09	44.84
Centro Internacional de Mejoramiento del Maíz y Trigo	CIMMYT	1182	12.44	1.93	60.71	3.64	23.01	82.74	39.09
Instituto Nacional de Cancerologia	INCAN	1179	12.28	2.05	53.71	1.19	8.40	37.15	27.99
Universidad Autonoma de Sinaloa UAS	UAS	1154	4.17	0.95	31.52	0.87	8.58	43.93	34.23
Instituto Nacional de Astrofisica, Optica y Electronica	INAOE	1090	8.52	1.13	48.57	1.47	7.80	57.16	46.51
Centro de Investigacion en Alimentacion y Desarrollo	CIAD	965	7.15	0.72	32.18	0.21	4.97	29.74	48.81
El Colegio de la Frontera Sur	ECOSUR	955	6.59	0.93	28.18	1.88	5.76	50.26	51.94
National Institute of Cardiology - Mexico	INC	912	8.44	1.22	46.03	1.32	8.55	31.14	36.84
Instituto Potosino Investigacion Cientifica y Tecnologica	IPICT	883	9.60	0.87	40.02	0.57	6.23	40.54	49.72
Hospital Infantil de Mexico Federico Gomez	HIMFG	839	10.76	0.72	34.54	0.72	6.44	38.14	22.53
Universidad de Colima	U de C	708	3.71	0.64	30.95	0.56	4.66	41.38	42.23
Universidad Juarez Autonoma de Tabasco	UJAT	690	4.73	0.48	21.09	0.00	1.74	25.07	42.46
Instituto Nacional de Psiquiatria Ramon de la Fuente Muniz	INPRFR	653	7.10	3.99	41.47	6.28	14.55	48.09	32.92
Universidad Autonoma de Zacatecas	UAZ	650	3.17	0.50	20.31	0.15	2.62	35.23	43.08

Centro de Investigacion Cientifica de Yucatan	CICY	645	6.45	0.75	33.28	1.09	4.19	38.60	56.90
Universidad Autonoma de Ciudad Juarez	UACJ	627	2.29	0.86	27.29	0.64	3.99	45.61	44.82
Universidad Autonoma de Coahuila	UAdeC	623	3.54	0.68	34.91	0.48	4.17	32.58	40.13
Instituto Nacional de Medicina Genomica	INMEGEN	588	12.25	1.32	45.39	0.85	9.18	31.12	29.42
Centro de Investigaciones en Optica A.C.	CIO	579	7.52	0.78	31.70	0.35	5.18	41.45	52.33
Instituto Mexicano del Petroleo	IMP	553	4.46	0.54	36.95	0.00	3.25	24.41	52.26
Universidad Autonoma de Tamaulipas	UAT	534	3.30	0.70	16.44	0.94	2.43	36.52	41.76
Universidad Juarez del Estado de Durango	UJED	507	4.61	0.58	24.54	0.39	3.75	39.64	39.84
Centro de Investigacion en Materiales Avanzados S.C.	CIMAV	457	5.57	0.51	32.30	0.00	1.53	34.35	40.48
Universidad Autonoma de Aguascalientes	UAA	452	3.53	0.69	25.28	0.88	5.75	34.96	45.35
Centro de Investigacion Quimica Aplicada	CIQA	416	5.07	0.62	32.60	0.00	4.57	34.86	45.91
Universidad Americas Puebla	UDLAP	403	4.11	0.95	36.95	0.50	5.96	45.16	39.95
Hospital General de Mexico	HGM	398	7.80	1.36	44.81	3.02	10.30	39.70	18.59
Centro de Investigacion y Docencia Economicas	CIDE	391	3.37	0.71	34.72	0.51	5.88	47.83	64.19
Instituto Tecnologico Autonomo de Mexico	ITAM	356	3.79	0.86	35.90	0.28	5.62	60.39	51.97
CIMAT - Centro de Investigacion en Matematicas	CIMAT	328	3.49	0.60	26.77	0.00	3.66	50.00	51.83
Universidad Panamericana - Ciudad de Mexico	UP	311	3.05	0.76	41.37	0.32	5.79	54.02	41.80
Universidad Autonoma de Chiapas	UNACH	297	2.54	1.34	40.14	1.68	9.09	48.15	29.29
Universidad de Monterrey	UDEM	239	3.68	0.66	31.90	0.00	5.02	59.41	26.78
Colegio de Mexico	COLMEX	227	1.20	0.60	25.57	0.44	3.96	29.52	62.11
Universidad Anahuac	UA	224	2.87	0.46	23.18	0.00	2.23	27.68	38.39
Centro de Investigacion & Desarrollo Tecnologico en Electroquimica S.C.	CIDETEQ	207	5.59	0.72	44.88	0.00	4.35	35.75	56.04

Anexo 7: Perfil cuantitativo de las 22 áreas de investigación mexicana en el WoS para el año 2017

Áreas ESI	Share	CNCI	% Top 1%	% Top 10%	% CI	% Q1
Agricultural Sciences	5.4030782	0.9219532	1.0956175	7.3705179	38.247012	33.165323
Biology & Biochemistry	5.3169734	0.7908587	0.5060729	7.2874494	38.663968	40.653728
Chemistry	8.7181143	0.7850925	0.5555556	4.4444444	41.296296	34.189723
Clinical Medicine	15.202885	1.3299004	1.7699115	8.1415929	43.079646	46.971346
Computer Science	1.4799268	0.6607724	0	6.5454545	46.909091	34.814815
Economics & Business	1.0278764	0.6458351	0	6.2827225	57.591623	26.344086
Engineering	7.7924874	0.761524	0.5524862	5.6629834	41.643646	40.86774
Environment/Ecology	6.3771392	0.7938397	1.0126582	5.907173	45.822785	30.119454
Geosciences	2.7930255	0.870179	0.7707129	5.3949904	58.959538	37.790698
Immunology	2.2602519	0.9125305	1.1904762	8.3333333	41.666667	33.014354
Materials Science	4.3321494	0.5269467	0.1242236	2.9813665	47.329193	38.548185
Mathematics	2.3947907	0.655107	0.2247191	4.0449438	47.41573	24.379233
Microbiology	1.7113335	0.7941365	0.3144654	1.2578616	48.427673	19.426752
Molecular Biology & Genetics	2.3248305	1.0028155	1.6203704	8.7962963	54.398148	40.97561
Multidisciplinary	0.1022495	2.8610105	5.2631579	26.315789	47.368421	84.210526
Neuroscience & Behavior	2.7553546	0.8646477	1.171875	4.296875	44.921875	39.285714
Pharmacology & Toxicology	2.9006565	0.842508	0.5565863	6.864564	43.413729	40.712946
Physics	7.1789904	0.9126121	1.1244378	9.1454273	53.973013	39.548872
Plant & Animal Science	11.048326	0.8763755	0.9254749	6.8679981	44.76376	29.26347
Psychiatry/Psychology	1.6844258	0.9808141	1.9169329	7.6677316	63.578275	32.792208
Social Sciences, general	4.8380153	0.6691253	0.3337041	4.4493882	42.269188	28.121354
Space Science	2.3571198	1.6421196	3.196347	14.611872	86.52968	78.26087

Share: Fracción de la producción científica mexicana

CNCI: *Category Normalized Citation Impact*

%Top 1%: Porcentaje de la producción científica mexicana en el 1% de lo más citado

% Top 10%: Porcentaje de la producción científica mexicana en el 10% de lo más citado

% CI: Porcentaje de documentos en colaboración Internacional

%Q1: Porcentaje de documentos publicados en revistas de cuartil 1

Anexo 8: Perfil de desempeño temático de las 22 áreas de investigación mexicana en el WoS para el año 2018

Áreas ESI	Share	CNCI	% Top 1%	% Top 10%	% International Collaborations	% Q1
Agricultural Sciences	5.7285621	0.7795833	0.5286344	5.9030837	38.590308	32.47331
Biology & Biochemistry	4.0478474	0.7210817	0.3740648	5.4862843	45.63591	31.841432
Chemistry	8.49947	0.8041205	0.415677	3.8004751	40.320665	34.243176
Clinical Medicine	15.464594	1.4059909	2.0234987	7.9634465	43.603133	44.448127
Computer Science	1.978499	0.6767173	0	4.3367347	44.132653	28.051948
Economics & Business	0.9640135	0.7435199	0	7.8534031	52.356021	29.946524
Engineering	7.7726745	0.7006003	0.5194805	5.3246753	42.012987	35.473197
Environment/Ecology	6.6925756	0.7116881	0.7541478	5.8823529	49.245852	30.927052
Geosciences	2.9172765	0.7635484	0.3460208	6.0553633	60.726644	37.279152
Immunology	2.1601978	0.9091759	1.1682243	7.4766355	51.869159	28.571429
Materials Science	4.5323777	0.4738589	0.1113586	1.55902	43.76392	31.715576
Mathematics	2.2611417	0.7534828	0.4464286	6.0267857	50	26.818182
Microbiology	1.8422248	0.7193616	0.2739726	3.2876712	45.205479	28.846154
Molecular Biology & Genetics	2.1349619	1.0895284	1.1820331	6.6193853	46.335697	25.653207
Multidisciplinary	0.1463685	2.9982103	6.8965517	13.793103	58.62069	39.285714
Neuroscience & Behavior	2.558926	1.102217	0.591716	6.1143984	47.140039	35.728543
Pharmacology & Toxicology	2.4680765	0.8149695	0.204499	6.7484663	41.513292	45.473251
Physics	7.1114925	1.1158289	1.2775018	11.213627	54.932576	37.962963
Plant & Animal Science	11.664059	0.7676598	0.692341	4.9762008	46.300303	26.016616
Psychiatry/Psychology	1.8270832	0.9044108	1.9337017	6.9060773	51.381215	43.055556
Social Sciences, general	5.1329935	0.7385812	0.4916421	5.8013766	41.986234	26.485398
Space Science	2.0945844	1.1911959	1.686747	12.771084	88.192771	81.686747

Share: Fracción de la producción científica mexicana

CNCI: *Category Normalized Citation Impact*

%Top 1%: Porcentaje de la producción científica mexicana en el 1% de lo más citado

% Top 10%: Porcentaje de la producción científica mexicana en el 10% de lo más citado

% CI: Porcentaje de documentos en colaboración Internacional

%Q1: Porcentaje de documentos publicados en revistas de cuartil 1

Anexo 9: Perfil de desempeño temático de las 22 áreas de investigación mexicana en el WoS para el año 2019

Áreas ESI	Share	CNCI	% Top 1%	% Top 10%	% CI	% Q1
Agricultural Sciences	5.800768	0.7759879	0.3115265	5.2959502	38.707165	37.294762
Biology & Biochemistry	4.3731647	0.7732596	0.9297521	6.1983471	44.008264	33.966245
Chemistry	8.7869889	0.6149254	0.1542416	3.7017995	40.051414	31.372549
Clinical Medicine	15.035012	1.6868544	2.4338942	9.3149038	44.140625	46.378576
Computer Science	1.8341992	0.7476613	0.2463054	6.4039409	50.246305	29.776675
Economics & Business	1.2197877	0.8182807	1.1111111	6.6666667	58.888889	19.702602
Engineering	7.9241021	0.7518786	0.2280502	5.5872292	40.763968	34.8
Environment/Ecology	6.627513	0.8769008	1.2269939	7.4982958	47.989093	34.866163
Geosciences	2.8777953	0.7248656	0.3139717	4.0816327	64.207221	29.245283
Immunology	2.2633838	0.6857537	0.7984032	5.3892216	42.315369	27.744511
Materials Science	4.7933138	0.5385208	0	2.4505184	40.904807	35.096154
Mathematics	2.394398	0.5852847	0.3773585	3.9622642	52.641509	24.497992
Microbiology	1.9426248	0.8411402	0.4651163	6.2790698	51.860465	13.679245
Molecular Biology & Genetics	2.2498306	0.7896331	0.6024096	6.2248996	47.791165	22.474227
Multidisciplinary	0.1400497	1.9573032	3.2258065	19.354839	58.064516	38.709677
Neuroscience & Behavior	2.9455613	0.988577	0.6134969	4.2944785	41.411043	37.941628
Pharmacology & Toxicology	2.3356675	0.7201174	0.1934236	4.0618956	37.524178	40.236686
Physics	6.6094421	1.0783631	1.2303486	9.8427888	55.365687	40.013727
Plant & Animal Science	10.838039	0.8352762	0.8336807	5.6273447	45.102126	31.530343
Psychiatry/Psychology	2.5660718	0.5975519	0.528169	4.2253521	33.978873	14.690265
Social Sciences, general	4.4093065	0.643249	0.307377	5.0204918	43.954918	25.56701
Space Science	2.0329794	1.5740807	2.2222222	12.666667	86.666667	83.111111

Share: Fracción de la producción científica mexicana

CNCI: *Category Normalized Citation Impact*

%Top 1%: Porcentaje de la producción científica mexicana en el 1% de lo más citado

% Top 10%: Porcentaje de la producción científica mexicana en el 10% de lo más citado

% CI: Porcentaje de documentos en colaboración Internacional

%Q1: Porcentaje de documentos publicados en revistas de cuartil 1

Anexo 10: Valores anuales de los indicadores que caracterizan el perfil  
 cienciométrico de México en WoS entre 2000-2019.

NSP: Índice Nacional de Productividad

%Q1: Porcentaje de documentos publicados en revistas de cuartil 1

%Top 1%: Porcentaje de la producción científica mexicana en el 1% de lo más  
 citado

% Top 10%: Porcentaje de la producción científica mexicana en el 10% de lo más  
 citado

CNCI: *Category Normalized Citation Impact*

% CI: Porcentaje de documentos en colaboración Internacional

CINV: El coeficiente de inventiva se refiere a las patentes solicitadas por los  
 residentes cada 10, 000 habitantes

Años	NSP	%Q1	% Top 1%	% Top 10%	CNCI	%CI	CINV
2000	0.89202978	38.7703153	0.68016695	6.708919462	0.76080309	35.1522646	0.43579441
2001	0.95553174	37.1482824	0.476590973	6.980656013	0.76100285	35.9686011	0.5324126
2002	0.91918184	40.5119454	0.705563094	6.919945726	0.74450389	36.0515604	0.517285
2003	0.89205348	41.850357	0.389958567	6.982695589	0.71542392	35.3765537	0.45401181
2004	1.00608762	41.6842698	0.486223663	6.515397083	0.72129299	35.289033	0.54059261
2005	0.94209442	40.7858182	0.458381081	6.834045213	0.74516287	35.8891551	0.55091635
2006	0.95505721	40.5143073	0.699635806	7.063446425	0.79276391	39.6875599	0.53365487
2007	0.97947851	43.233041	0.673732643	7.419275327	0.76093777	37.7043793	0.58715494
2008	0.97462231	39.7734559	0.748320098	7.170128283	0.78257114	39.1722663	0.61814585
2009	0.91132633	38.4163347	0.806451613	7.039493518	0.79689662	38.5966235	0.73090129
2010	0.90620761	39.1103302	0.891011476	7.156604177	0.76435898	39.204505	0.83353081
2011	0.88570737	39.0623633	0.956872394	7.279064999	0.80453315	40.243319	0.92052003
2012	0.89281648	39.0853561	0.943576666	7.121453618	0.86577597	39.6047179	1.10169201
2013	0.9241719	39.4081797	0.960468521	7.191800878	0.81244245	39.8945827	1.01912727
2014	0.90887815	39.3854357	0.968428124	7.215629198	0.85634847	41.1609942	1.03360781
2015	0.86527133	37.3058981	1.067615658	6.756408273	0.88379822	41.4771262	1.11933325
2016	0.9020059	38.8879113	0.933814681	7.239470517	0.9032073	43.1191336	1.06216179
2017	0.8824984	37.5041177	0.989785772	7.032450511	0.83776263	44.0658049	1.06910451
2018	0.86312828	35.1301496	0.921298509	6.383903351	0.82673231	44.5526053	1.23226111
2019	0.85359367	35.0954885	0.926457326	6.358120849	0.85991577	44.6585226	1.03414839

