

CIENCIOMETRÍA: FACTORES DE IMPACTO Y MAPAS DE LA CIENCIA



MARCELO MENDOZA

Académico del Departamento de Informática de la Universidad Técnica Federico Santa María (UTFSM) desde 2011. Doctor en Ciencias mención Computación, Universidad de Chile (2007); Magíster en Ciencias de la Ingeniería Informática, UTFSM (2001) e Ingeniero Civil Electrónico, UTFSM (2001). Director del Programa de Magister en Ciencias de la Ingeniería Informática (UTFSM) desde 2014 y Presidente de la Asociación Chilena de Reconocimiento de Patrones (ACHIRP) desde 2015. Líneas de investigación: Minería de Texto, Recuperación de Información en Texto, Minería de Datos en Redes Sociales. mmendoza@inf.ut fsm.cl



MIGUEL GUEVARA

Académico del Departamento de Computación e Informática, Facultad de Ingeniería de la Universidad de Playa Ancha (UPLA) desde 2008. Doctor en Ingeniería Informática, Universidad Técnica Federico Santa María, UTFSM (2016); Magíster en Ciencias de la Ingeniería Informática, UTFSM (2013); Magíster en Tecnologías de la Información, UTFSM (2007) e Ingeniero en Informática, Universidad Central del Ecuador (2001). Coordinador del Núcleo de Innovación (UPLA). Líneas de investigación: Minería de Datos Científicos desde la Web, Cienciometría, Redes Complejas, Educación y Tecnología. miguel.guevara@upla.cl

La cienciometría es la disciplina que analiza la producción científica como objeto de estudio. Son sus principales áreas la medición del impacto científico, el estudio de patrones de colaboración, el análisis de modelos de propagación de información y la construcción de mapas de la ciencia.

Existe un amplio consenso en cienciometría en cuanto a que el número de artículos publicados por autor no refleja el impacto que éste tiene en su comunidad. Gupta *et al.* [1] mostraron que la distribución de citas por artículos sigue una *power-law*, es decir unos pocos artículos realmente logran tener impacto, mientras que la mayoría de ellos registran pocas citas y su impacto es prácticamente nulo. Por ello, contar artículos es *naïve* en términos de impacto.

FACTORES DE IMPACTO

Entendemos por impacto científico a la contribución específica de un artículo en un área. Un artículo de alto impacto necesariamente es referenciado por otros artículos. Debido a lo anterior, la señal predominante del impacto de un artículo se refleja en las citas.

Tanto Web of Science¹ (WoS) como Scopus², los dos principales servicios de indexación de artículos científicos, han incorporado factores de impacto como parte de sus servicios. WoS usa el Journal Impact Factor (IF), definido como el cociente entre el número de citas en un año y el número de artículos publicados por la revista los dos años anteriores. Se liberan anualmente en el Journal Citation Report (JCR).

1 <http://webofknowledge.com>
2 <http://www.scopus.com>

En Chile, la asignación de recursos del Estado para financiar investigación a través de Conicyt ha utilizado principalmente JCR como indicador de impacto de los investigadores. Se ha asumido que la revista es un *proxy* del impacto futuro del artículo. El supuesto es que un artículo publicado en una revista con alto IF será un artículo con muchas citas. Sin embargo, en ciencia métrica existe consenso en que la correlación entre IF y la mediana de citas por artículo es débil [2], lo que cuestiona la capacidad predictiva de IF como indicador de impacto.

A nivel de autor, actualmente es el H-index [3] el índice de impacto más usado. Un H-index igual a h indica que los h artículos más citados de un autor tienen más de h citas. El H-index es eficaz en descartar artículos que pertenecen al long-tail de la distribución de citas por artículos. Una variante del H-index, denominada G-index [4], busca además medir el efecto del número de citas sobre el umbral h . Un G-index igual a g indica que los g artículos más citados de un autor tienen más de g citas en promedio. Esto favorece a autores cuyos artículos más citados tienen muchas citas, como muestra la Figura 1. Se puede mostrar que $H\text{-index} \leq G\text{-index}$.

Scopus ha incorporado nuevos índices de impacto a su servicio de indexación. Usa el Impact per Publication (IPP) que es equivalente al IF pero sobre una ventana de tres años. Además incorpora el Source Normalized IP [6] (SNIP) que introduce un factor de normalización según el largo de las listas de referencias (mientras mayor es el largo de la lista de referencias, menor es el valor de sus citas), y el SJR (Scimago Journal Rank) [7] que corresponde al PageRank calculado en la red de citaciones entre revistas.

La pregunta acerca de cuál servicio de indexación es mejor es controversial. Sin embargo, existe evidencia en favor de Scopus en términos de mejor cobertura. Al año 2016, Scopus indexa 29,566 revistas, mientras que Web of Science registra 18,843 [5], con una intersección entre ambos servicios de 14,391 revistas. En términos de factores de impacto, Scopus ha incorporado IPP, SNIP y SJR los cuales son índices que presentan mejoras frente a IF [6,7].

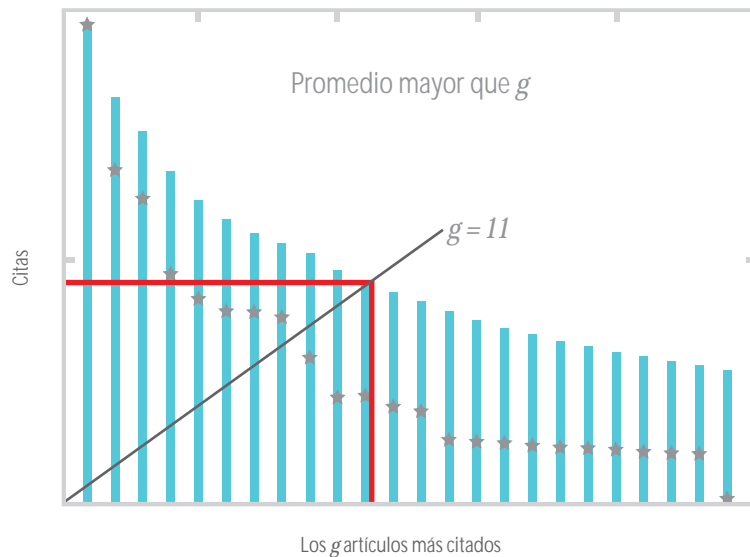


FIGURA 1. UN AUTOR OBTIENE G-INDEX 11 DEBIDO A QUE SUS ARTÍCULOS MÁS CITADOS TIENEN MUCHAS CITAS, LO CUAL SUBE EL PROMEDIO DE CITAS SOBRE G. EN EL EJEMPLO, EL MISMO AUTOR OBTENDRÍA UN H-INDEX 8.

LOS MAPAS DE LA CIENCIA

Los factores de impacto son incapaces de producir una adecuada caracterización a un nivel de granularidad mayor. Por ejemplo, si necesitamos evaluar un grupo de investigadores, una institución o incluso un país, ningún factor de impacto es capaz por sí sólo de representar adecuadamente la productividad científica. Esto se debe a que al aumentar la granularidad se incorpora una compleja red de interacciones y patrones de producción, incapaces de ser representados adecuadamente por un solo indicador.

Desde la ciencia métrica se ha abordado el problema de la caracterización a mayores niveles de granularidad usando mapas, redes de información en las que los nodos representan áreas de la ciencia (en algunos casos también revistas científicas) y los enlaces representan similitud entre estas áreas. Un compendio de la hetero-

génea lista de mapas realizados para caracterizar el conocimiento científico fue realizado por Katy Börner en el denominado Atlas de la Ciencia [8]. La mayoría de estos mapas han sido realizados por expertos (bibliómetros) y están basados en patrones de citas entre revistas o artículos científicos.

Pero las redes de información no son exclusivas del área de la ciencia métrica. El año 2011 el grupo Collective Learning de MIT Media Lab, dirigido por el chileno César Hidalgo, en colaboración con Ricardo Hausmann de la Universidad de Harvard, publicaron el Atlas de la Complejidad Económica [9]. En el Atlas y artículos relacionados, los autores pretenden caracterizar la complejidad de la matriz productiva económica a nivel global. La naturaleza del problema era similar. Era necesario comparar la producción económica a altos niveles de granularidad (países) sobre unidades de distinta complejidad y tamaño. Nos propusimos abordar con Hidalgo el tema de la producción científica desde la base que nos ofrecía el Atlas de la Complejidad Económica.

ANÁLISIS DE VENTAJAS COMPARATIVAS

Economías de tamaños distintos son comparables si se evalúan en términos relativos. Una ventaja comparativa es la habilidad de elaborar un producto a un costo marginal sobre otro producto. Esto se debe a que la producción económica está condicionada a la estructura de la matriz productiva. El desarrollo de productos se realiza por áreas de producción, debido a las facilidades y capacidades desarrolladas en esa área. En ciencia ocurre algo similar. Son las capacidades de los investigadores y las

facilidades de las instituciones que los acogen las que explican el desarrollo del nuevo conocimiento.

Es posible caracterizar la matriz productiva usando ventajas comparativas reveladas (RCA) [10] a partir de su producción actual. El índice RCA de un productor c en un producto f se define como el cociente entre la producción relativa de c en f a su producción total, sobre la producción global relativa en f a la producción global total. Por ejemplo, en Ciencia de la Computación se produjeron 1,222,228 artículos científicos entre los años 1996 y 2011 (fuente SCImago³), lo que representa el 4% de la producción científica global. Hong Kong produjo 174,400 artículos, de los cuales 16,684 corresponden a Ciencia de la Computación, lo que representa el 9.5%

de su producción científica. Luego, en términos de número de artículos, Hong Kong tiene una $RCA = 2.37$ en Ciencia de la Computación. En el área de cienciometría, la medida RCA se conoce como Índice de Actividad (AI).

Se dice que c tiene una ventaja comparativa para producir f si su índice RCA es mayor que 1. Notar que podemos reemplazar el índice de producción por citas o H-index. Además, el análisis basado en productividad nos permite caracterizar la diversidad de la matriz productiva y la ubicuidad de la producción en una determinada área. La diversidad productiva de c consiste en el número de áreas en las cuales c tiene un $RCA > 1$. La ubicuidad de un área f consiste en el número de productores donde el área tiene un $RCA > 1$. Hemos estudiado los resultados del análisis de diversidad y ubicuidad productiva para la ciencia que realizamos usando datos de SCImago⁴ [11]. Las Tablas 1 y 2 muestran resultados de ubicuidad y diversidad para producción de artículos, citas y H-index, respectivamente.

La Tabla 1 muestra que las áreas con mayor ubicuidad son "Agricultura y Biología" y "Medioambiente". "Ciencia de la Computación" es una de las áreas menos ubicuas en los tres indicadores. La Tabla 2 muestra los países con mayor diversidad en su matriz productiva. Se pueden observar algunas diferencias al variar el índice de productividad. Por ejemplo, en producción de artículos, Canadá y Holanda ocupan el primer lugar, con RCA en 16 áreas, mientras que en producción de citas la matriz de mayor diversidad la tiene Malasia y en H-index los Emiratos Árabes Unidos. Chile ocupa el lugar 35 en la tabla, ordenada por diversidad en producción de artículos.

Nuestro *package* para R, *diverse* [12], facilita el cálculo de las medidas de RCA y ubicuidad, además de varias medidas de diversidad. En [13] describimos el detalle de este tipo de redes y también construimos mapas de la ciencia con datos de ejemplo incluidos en el *package*.

Research field	Ubi Articles	Ubi Citations	Ubi H-index
Agricultural and Biological Sciences	178	179	165
Arts and Humanities	76	54	79
Biochemistry, Genetics and Molecular Biology	24	18	87
Business, Management and Accounting	67	45	68
Chemical Engineering	53	68	74
Chemistry	55	60	81
Computer Science	37	44	64
Decision Science	54	57	55
Dentistry	67	64	65
Earth and Planetary Science	124	122	128
Economics, Econometrics and Finance	82	52	79
Energy	84	98	96
Engineering	38	58	83
Environmental Science	172	166	153
Health Professions	53	32	63
Immunology and Microbiology	141	130	120
Materials Science	42	60	72
Mathematics	72	79	80
Medicine	124	102	141
Multidisciplinary	84	34	47
Neuroscience	34	24	59
Nursing	75	46	69
Pharmacology, Toxicology and Pharmaceutics	70	72	97
Physics and Astronomy	53	59	71
Psychology	40	27	59
Social Sciences	126	115	139
Veterinary	126	130	113

TABLA 1. ÍNDICES DE UBICUIDAD PARA PRODUCCIÓN DE ARTÍCULOS, CITAS Y H-INDEX.

3 <http://www.scimagojr.com/>

4. Datos de Scimago 1996-2011, correspondiente a la producción en 238 países en 27 áreas. Se registran 29,895,499 artículos con 429,922,232 citas.

Country	Div Articles	Div Citations	Div H-index
Canada	16	14	13
Netherlands	16	14	13
United States	15	13	12
Israel	15	12	15
Norway	15	12	14
Australia	15	12	12
United Kingdom	14	15	12
Ireland	14	10	14
Finland	14	8	14
United Arab Emirates	13	14	17
New Zealand	13	12	15
Switzerland	13	12	13
Belgium	12	13	14
Malaysia	12	16	14
Sweden	12	11	11
Portugal	11	14	14
Thailand	11	13	15
Saudi Arabia	11	13	13
India	11	13	10
South Korea	11	12	14
Indonesia	11	12	13
Croatia	11	12	11
Mexico	11	11	16
Serbia	11	11	14
Singapore	11	11	13
China	11	11	10
Qatar	11	10	12
Argentina	11	10	12
Denmark	11	9	12
Austria	11	7	9
Hong Kong	10	15	15
Spain	10	11	15
Czech Republic	10	11	12
Italy	10	10	13
Chile	10	10	10

TABLA 2.
ÍNDICE DE DIVERSIDAD PARA PRODUCCIÓN DE ARTÍCULOS, CITAS Y H-INDEX.

EL RESEARCH SPACE

Deseábamos construir una red aplicable no solo a países sino también a instituciones y personas. Por lo que decidimos avanzar nuestra investigación hacia la construcción de un mapa de la ciencia que permitiera predecir la producción científica. Denominamos a este mapa, el Espacio de Investigación (*Research Space*) [14].

La mayoría de mapas de la ciencia, como los mapas recogidos en el libro de Katy Börner,

construyen las relaciones entre áreas basados en patrones de citas. Ésta es una excelente forma de *ordenar* la ciencia. Nuestro interés era distinto, ya que nos interesaba capturar las capacidades productivas de los generadores de nuevo conocimiento para detectar cómo se relacionan las áreas de investigación desde esta mirada *productiva*. Para esto, requerimos trabajar al mayor nivel de profundidad productiva, esto es, a nivel de individuos. Aquí afrontamos un primer desafío, y es que las bases de datos científicas disponibles (en 2013) a nivel de usuario eran inexistentes o contenían serios problemas de ambigüedad de nombres. Resolvimos este

problema descargando datos de Google Scholar⁵, la que puede ser vista como una base de datos curada por los mismos usuarios (una especie de *crowdsourcing*). Esto se debe a que los usuarios registrados deben indicar explícitamente cuáles publicaciones son de su autoría.

Con técnicas de *web scraping* y por un período de más de dos meses, logramos obtener 319,049 usuarios y 475,774 publicaciones que vinculamos a 16,873 revistas científicas.

Con este conjunto de datos conectamos cada publicación –y en consecuencia cada autor– a una categoría de la ciencia en la clasificación de SCImago (la que incluye 308 categorías agrupadas en 27 áreas). Hicimos lo propio utilizando la clasificación UCSD, que fue propuesta por uno de los mapas de citas más difundidos y que es el mapa con el que comparamos nuestros resultados. Esta clasificación incluye 554 categorías agrupadas en 13 disciplinas.

Para definir los valores de similaridad (cercanía) entre categorías de la ciencia, calculamos la probabilidad condicional de que un autor publicara en dos áreas al mismo tiempo. Es decir, cada enlace representa similaridad entre áreas de la ciencia considerando capacidades productivas demostradas en artículos publicados. Nótese aquí la diferencia con mapas de la ciencia basados en citas, debido a que el hecho de citar un artículo de cierta área, no significa necesariamente que el autor tenga las capacidades de publicar en esa área. Para tomar el ejemplo de un artículo de “Redes Neuronales Artificiales” (un área de fuerte desarrollo en Computación), el autor de ese artículo puede citar artículos en el área de la “Fisiología del Cerebro”, pero no necesariamente es capaz de publicar en esa área. En sentido opuesto, un artículo en el área de la “Fisiología del Cerebro” puede citar un artículo de algún método computacional, pero esto no implica que los autores sean capaces de aportar al conocimiento nuevo en el área de la Computación. Nuestro mapa propone esa diferencia metodológica en la construcción de los enlaces.

5. <http://scholar.google.com/>

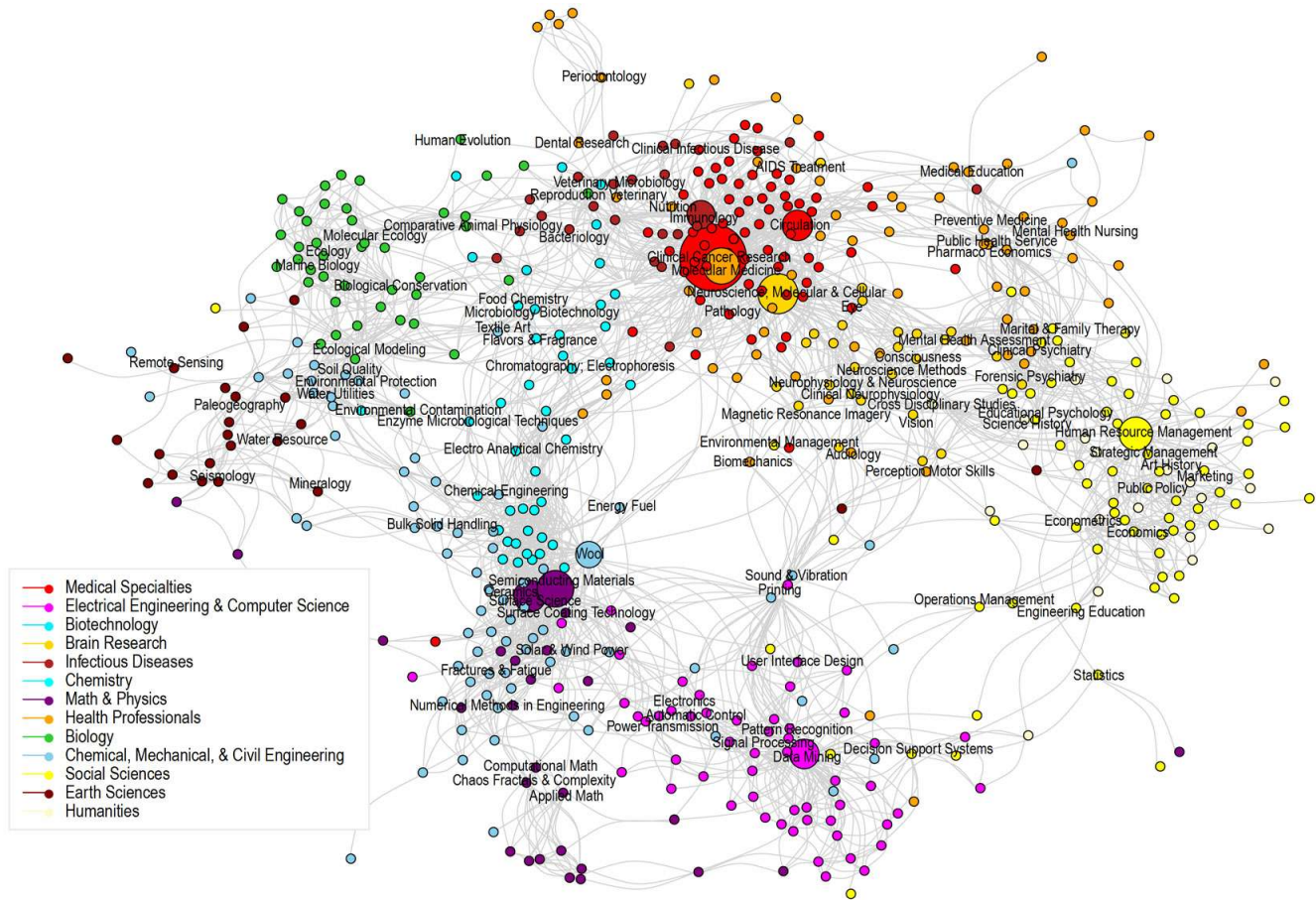


FIGURA 2.

EL ESPACIO INVESTIGACIÓN. LOS NODOS REPRESENTAN CATEGORÍAS EN CLASIFICACIÓN UCSD. LOS ENLACES ESTÁN DEFINIDOS EN BASE A LA PROBABILIDAD CONDICIONAL DE QUE UN AUTOR PUBLIQUE EN LAS DOS ÁREAS AL MISMO TIEMPO. COLORES DEFINIDOS POR DISCIPLINA.

En la Figura 2, presentamos el Espacio Investigación en clasificación UCSD. En esta Figura, utilizando una analogía geográfica, se aprecia por ejemplo, computación en el Sur, "Medicina" en el Norte, "Ciencias Sociales" en el Este y "Ciencias de la Tierra" en el Oeste. Las "Ciencias Básicas" en el Suroeste, "Biología" en el Noroeste y las "Ciencias de la Salud" en el Noreste. Se pueden apreciar conexiones que hacen mucho sentido, por ejemplo, la cercanía entre categorías de la "Matemática" y categorías de la "Computación".

Una vez construido el mapa, podemos posicionar a los productores sobre él. Los productores son entidades a distintos niveles de agrupación, por ejemplo, individuos, instituciones o países.

Para visualizar la producción podemos definir estados basados en alguna medida, como por ejemplo los porcentajes de producción en cada categoría, o, en el caso de países, definir estos estados según sus valores de "ventajas comparativas" (RCA). El primero (porcentajes) es un análisis local o interno, mientras que el segundo es un análisis en relación con los demás productores de ciencia.

Para nuestro análisis definimos dos estados básicos: Inactivo y Activo. A su vez, el estado Activo lo podemos subdividir en Naciente, Intermedio y Desarrollado. En el caso de porcentajes de producción, estos tres últimos estados se pueden definir utilizando cuartiles, mientras que en el caso de RCA estos estados se definen

como Desarrollado para valores mayores a 1 (con ventajas comparativas), Intermedio para valores de RCA entre 0.5 y 1, y Naciente para valores RCA menores a 0.5.

Con los datos de producción que obtuvimos de Google Scholar, y asociando cada usuario a una institución y cada institución a un país, podemos visualizar y evaluar la producción de individuos, instituciones y países. Por ejemplo, en la Figura 3A presentamos el espacio investigación del Instituto Tecnológico de Massachusetts MIT utilizando porcentajes de producción para definir los estados; mientras que en la Figura 3B presentamos el espacio investigación de Chile, utilizando RCA para definir los estados. En el caso de MIT, se notan

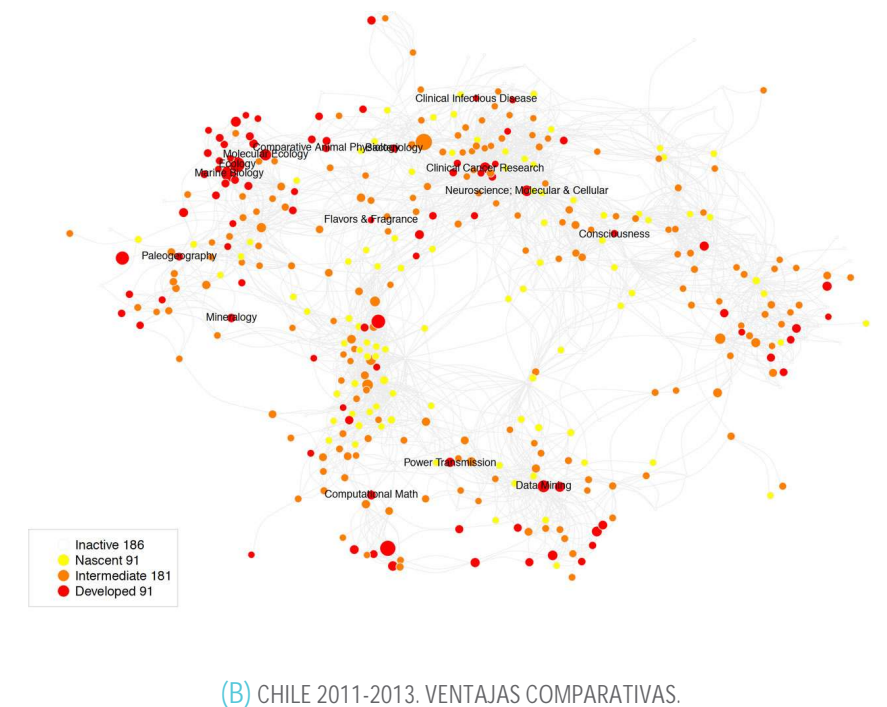
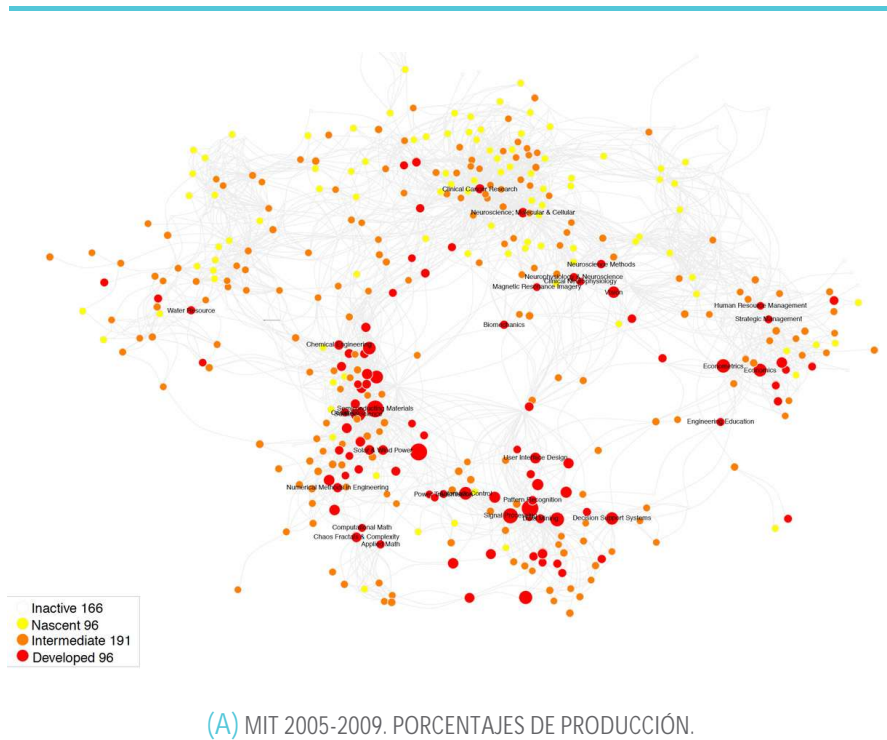


FIGURA 3. MAPAS SUPERPUESTOS DE EJEMPLO PARA INSTITUCIONES Y PARA PAÍSES. LOS COLORES INDICAN EL ESTADO DE DESARROLLO SEGÚN LA MEDIDA EVALUADA.

altos niveles de producción en áreas de la “Ingeniería”, la “Computación”, las “Ciencias Básicas” o la “Economía”. En el caso de Chile, para el período 2011-2013 se aprecian 91 áreas Desarrolladas ($RCA > 1$) entre las que destacan “Investigación Clínica en Cáncer”, “Ecología”, “Biología Marina”, “Neurociencia”, “Minería de Datos” o “Matemática Computacional”.

Para evaluar las capacidades predictivas de nuestro mapa, utilizamos una medida de *densidad* activa que calcula qué tan conectado se encuentra un nodo inactivo con nodos ya activados. Para esto, para cada nodo inactivo en un tiempo t , calculamos el radio entre la suma de enlaces con nodos ya activos y lo dividimos por la suma total de enlaces conectados a ese nodo inactivo. Esta medida la calculamos tanto para la transición de Inactivo a Activo, como para las transiciones de Naciente a Desarrollado y de Intermedio a Desarrollado. Después, construimos una lista ordenada de áreas que se desarrollarán (que pasan a un estado superior) en el futuro y evaluamos esta lista con la datos reales del tiempo $t+1$. Conducimos la evaluación en dos intervalos de tiempo, entre 2008 y 2010 (t) y, entre 2011 y 2013 ($t+1$).

Evaluamos los resultados para cada transición y para cada tipo de productor, utilizando el área bajo la curva ROC (Receiver Operating Characteristic) que muestra en el eje Y la razón de verdaderos positivos (RTP) y en el eje X la razón de falsos positivos (RFP). Cada vez que un elemento Inactivo de la lista, ordenada según la medida de densidad, se Activa, la curva ROC avanza en el eje Y (RTP), en caso contrario, avanza en el eje X (RFP). El área bajo una curva (AUC) ROC obtenida aleatoriamente tiene valores cercanos a 0.5. Valores superiores a 0.6 se consideran aceptables, mientras que valores menores a 0.5 se consideran peores que *random*.

Consideramos solamente productores con suficientes datos para generar transiciones en los intervalos de tiempo estudiados, por lo que conducimos el análisis para 4850 individuos, 730 organizaciones y 77 países. En general para todos los productores y transiciones evaluadas

obtuvimos un promedio del área bajo la curva ROC mayor a 0.6, siendo el mejor caso el de individuos en transición Inactivo-Activo, con un valor en promedio de 0.89. En la misma transición, a nivel de instituciones obtuvimos un valor de 0.71 y a nivel de países un valor promedio de AUC igual a 0.68. En la transición de Naciente (nodos amarillos en Figura 3B) a Desarrollado (nodos rojos en Figura 3B), los resultados en

promedio de AUC fueron de 0.69 para instituciones y de 0.62 para países. Finalmente para la transición de Intermedio (nodos de color naranja en Figura 3b) a Desarrollado, obtuvimos valores en promedio de 0.64 para instituciones y de 0.65 para países. Nótese que no se evaluaron transiciones basadas en RCA para individuos debido a que la cantidad de transiciones a nivel de personas es insuficiente para conducir este análisis.

En [14] también mostramos que nuestro mapa no solo obtiene buenos resultados, sino que estos resultados son mejores (diferencia estadísticamente significativa) en comparación con el mapa UCSD [15] que es un mapa basado en patrones de citas. Esto a nivel de individuos y de instituciones, pero no así a nivel de países.

OBSERVACIONES FINALES

LOS RESULTADOS DE NUESTRA INVESTIGACIÓN PUEDEN APOYAR TAREAS DE POLÍTICA PÚBLICA Y POLÍTICA INSTITUCIONAL, POR EJEMPLO PARA PRIORIZAR ÁREAS Y LÍNEAS DE INVESTIGACIÓN EN UN PAÍS O UNA INSTITUCIÓN. ADEMÁS, EL ESPACIO INVESTIGACIÓN, TAMBIÉN PERMITIRÁ EVALUAR LA CALIDAD PRODUCTIVA CIENTÍFICA DE INDIVIDUOS, INSTITUCIONES Y PAÍSES DESDE UNA MIRADA MÁS COMPLEJA, VALE DECIR, CONSIDERANDO FACTORES COMO ESPECIALIZACIÓN/DIVERSIFICACIÓN O VENTAJAS COMPARATIVAS.

EN EL FUTURO, ESPERAMOS DIFUNDIR UNA HERRAMIENTA COMPUTACIONAL QUE PERMITA A LOS INDIVIDUOS, INSTITUCIONES Y PAÍSES, ANALIZAR SUS PROPIOS DATOS DE PRODUCCIÓN CIENTÍFICA, SUPERPONIENDO SU INFORMACIÓN AL ESPACIO INVESTIGACIÓN Y CALCULANDO DIFERENTES MEDIDAS DE DIVERSIDAD PRODUCTIVA COMO LAS PROPUESTAS EN [13]. ■

REFERENCIAS

- [1] Gupta, H., Campanha, J., Pesce, R. (2005). Power-law distributions for the citation index of scientific publications and scientists. *Brazilian Journal of Physics*, 35(4).
- [2] Eds. (2013). Beware the impact factor. *Nature Materials*, 12(89).
- [3] Hirsch, J. (2005). An index to quantify an individual's scientific research output. *PNAS* 102(46):16569-16572.
- [4] Egghe, L. (2006). Theory and practice of the g-index. *Scientometrics*, 69(1):131-152.
- [5] Mongeon, P., Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1):213-228.
- [6] Moed, H. (2010). Measuring contextual citation impact of scientific journals. *Journal of Informetrics*, 4(3):265-277.
- [7] González-Pereira, B., Guerrero-Bote, V., Moya-Aneón, F. (2010). A new approach to the metric of journals' scientific prestige: The SJR indicator. *Journal of Informetrics*, 4(3):379-391.
- [8] Borner, K. (2010). *Atlas of Science, visualizing what we know*. MIT Press.
- [9] Hausmann, R., Hidalgo, C., Bustos, S., Coscia, M., Chung, S., Jiménez, J., Simoes, A., Yildirim, M. (2011). *The Atlas of Economic Complexity*, MIT Press.
- [10] B. Balassa, "Trade Liberalisation and 'Revealed' Comparative Advantage", *The Manchester School*, vol. 33, n° 2, pp. 99-123, 1965.
- [11] Guevara, M., Mendoza, M. (2013). *Revealing comparative advantages in the backbone of Science*. CIKM, San Francisco, CA, USA.
- [12] M. R. Guevara, D. Hartmann, y M. Mendoza, diverse: Diversity Measures for Complex Systems. 2015. Disponible en <http://CRAN.R-project.org/package=diverse>
- [13] Guevara, M. R., Hartmann, D., & Mendoza, M. (2016). Diverse: an R Package to Measure Diversity in Complex Systems. *The R Journal*, 8(2), 60-78.
- [14] Guevara, M., Hartmann, D., Aristarán, M., Mendoza, M., Hidalgo, C. (2016). The Research Space: Using career paths to predict the evolution of the research output of individuals, institutions and nations. *Scientometrics*, 109(3):1695-1709.
- [15] K. Börner et al., "Design and Update of a Classification System: The UCSD Map of Science", *PLoS ONE*, vol. 7, n° 7, p. e39464, jul. 2012.