UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

Cartografía y análisis de redes de colaboración científica y redes sociales

Tesis Doctoral

GASTÓN PATRICIO OLIVARES FERNÁNDEZ

Ingeniero Civil Informático

Licenciado en Ciencias de la Ingeniería

Septiembre 2019

UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA
AGRONÓMICA, ALIMENTARIA Y DE BIOSISTEMAS

Cartografía y análisis de redes de colaboración científica y redes sociales

GASTÓN PATRICIO OLIVARES FERNÁNDEZ
Ingeniero Civil Informático
Licenciado en Ciencias de la Ingeniería

DIRECTORES

JUAN CARLOS LOSADA	JUAN PABLO CÁRDENAS
GONZALEZ	VILLALOBOS
Doctor en	Doctor en
Ciencias Físicas	Física de Sistemas Complejos

Septiembre 2019

*Para tí,
intrépido lector.*

Agradecimientos

Desde el año 2011 tuve la posibilidad de conocer a Juan Pablo Cárdenas y fue él quien me mostró el maravilloso mundo de la complejidad. Ya han pasado casi diez años desde entonces y hemos formado un sólido equipo, atravesando muchas aventuras a lo largo de este tiempo, llegando a tal punto que hoy es uno de mis directores de tesis. No puedo estar mas que agradecido contigo por la formación académica, humana e incluso laboral que me has brindado.

No puedo tampoco dejar de agradecer a Juan Carlos Losada, mi otro director de tesis que, desde que llegué a Madrid me recibió con los brazos abiertos y me ha ayudado (y soportado mis poco convencionales hábitos) en todo lo que he necesitado para que mi travesía por el viejo mundo sea mucho mas amena, de lo cual estaré eternamente agradecido.

También quiero agradecer a Rosa María Benito por toda la motivación que me ha brindado a lo largo de estos años que he estado viviendo en Madrid, sin contar la dedicación que le dedicó en las correcciones de mis trabajos y el haber podido contar con cualquier otra ayuda que va aún mas allá de lo académico.

Además, quiero agradecer a todos los integrantes el Grupo de Sistemas Complejos, con quienes he compartido a lo largo de estos años y a CONICYT PFCHA/DOCTORADO BECAS CHILE/2017 72180383 por haber financiado mi viaje y el programa de doctorado.

Tampoco puedo dejar de lado a mis padres de mis agradecimientos, quienes desde muy pequeño me han apoyado en cada una de mis excentricidades y a pesar de cualquier dificultad económica, siempre fui su prioridad.

Finalmente no puedo dejar de lado a Vivian, mi pareja con la cual hemos compartido mas de una década juntos, quien dejó de lado toda su vida en Chile para acompañarme por mi travesía en España, junto a Ivy, nuestra gata. El desarrollo de esta tesis no hubiese sido lo mismo sin la compañía física de ambas.

Resumen

En esta tesis se aborda el estudio de sistemas sociales desde la perspectiva de la teoría de redes complejas. En primer lugar se analizan redes de colaboración científica, que implican interacciones sociales, en teoría, presenciales y, posteriormente, se extiende el estudio a redes sociales *online*. En el caso de las redes de colaboración científica se han utilizado datos de la Web of Science para catorce países en un período de 2008-2018, mientras que en el caso de las redes sociales *online* se han analizado las conversaciones de Twitter durante la última semana de la segunda vuelta de las elecciones presidenciales de Chile el 2017.

En ambos casos se han modelizado los sistemas sociales como redes o grafos cuyos componentes poseen características propias cuantificables como son su Actividad en el sistema, la Respuesta del Sistema a esta actividad y su posición en la red.

En el caso de las redes de co-autoría, los autores corresponden a nodos del sistema mientras que sus enlaces representan las relaciones de co-autoría de cada publicación. En este caso hemos demostrado que las correlaciones más fuertes se dan entre el nivel de conectividad (coeficiente de grado) y la Actividad (número de publicaciones) de los autores y, con una correlación mayor, entre el Pagerank y la Actividad.

Una vez comprendido el funcionamiento global de una red de co-autoría, se procedió a realizar una transformación desde este tipo de red hacia un nuevo tipo llamado Red de Complemento Disciplinar, donde las disciplinas científicas asociadas a una publicación serán los nodos de este grafo enlazados a otros de acuerdo a las disciplinas científicas de la publicación. A través de la Actividad, Respuesta del Sistema y la Centralidad de cada disciplina se definió una métrica para medir la importancia de cada una de las disciplinas científicas. De estos resultados se obtiene que la disciplina *Biochemistry & Molecular Biology* es la más importante. Con esta medida de importancia se obtuvo un indicador en el intervalo -1 y 1 que permite medir y comparar la diversidad del conocimiento científico de los países analizados y demostrar que, en general, todos aquellos países definidos como desarrollados por el Banco Mundial poseen un valor mayor que cero de este indicador, mientras que para el resto de los países ese valor es menor a cero (negativo). Además

se encontró una fuerte correlación entre este indicador y el PIB per cápita de cada uno de los países.

Una vez conocidas las disciplinas más importantes, se diseñó un algoritmo para pronosticar el desarrollo de nuevas disciplinas científicas a nivel institucional, a través de una transformación de las redes de complemento disciplinar, denominada Cartografía del Conocimiento. Hemos comprobado que la posición (Centralidad) de las disciplinas no desarrolladas es más importante que la Actividad de sus disciplinas vecinas desarrolladas. Se cuantificó la bondad del algoritmo de predicción obteniendo una sensibilidad promedio de 0,84. Además de lo anterior, se desarrolló una herramienta web de análisis de las redes de co-autoría y de complemento disciplinar, que permite clasificar a los autores de acuerdo a su importancia y visualizar las relaciones entre ellos.

Finalmente, hemos extendido el estudio realizado en redes sociales de co-autoría (presenciales) a una red social *online*, midiendo la polarización política a través de un algoritmo basado en la posición de los usuarios en las redes de *Retweet* para la segunda ronda de las elecciones presidenciales chilenas del 2017. Partiendo de una pequeña cantidad de usuarios cuya opinión política se conoce de antemano, hemos estimado la opinión del resto de los usuarios que participan en la conversación. Con esto se observó que el día previo a la elección presidencial fue cuando el sistema alcanza su máximo valor de polarización y las opiniones de los usuarios de ese día se correlacionan bien con los resultados finales de las elecciones electorales.

Abstract

This thesis deals with the study of social systems from the complex network theory perspective. Firstly, scientific collaboration networks, which theoretically represent face-to-face relationships, were analyzed. Then, this study was extended to online social networks. In the case of scientific collaboration networks, the data for fourteen countries in the period of 2008-2018 has been obtained from Web of Science database. In the case of online social networks, Twitter conversations have been analyzed during the last week of the second round of the presidential elections of Chile in 2017.

In both cases, social systems have been modeled as networks or graphs whose components have their own quantifiable characteristics such as their Activity in the system, the System's Response to this activity and their position in the network.

In the case of co-author networks, the authors correspond to nodes of the system while their links represent the co-author relationships of each publication. In this case we have shown that the strongest correlations occur between the level of connectivity (degree coefficient) and the Activity (number of publications) of the authors and between Pagerank and author's Activity.

Once the overall operation of a co-authorship network was understood, a transformation was made from this type of network towards a new type called Disciplinary Complement Network, where the scientific disciplines associated with a publication are the nodes of this network linked to others according to the scientific disciplines of the publication. Through the Activity, System Response and Centrality of each discipline, a metric was defined to measure the importance of each of the scientific disciplines. The results obtained shown that the most important discipline is Biochemistry & Molecular Biology. With this measure of importance, a new metric was obtained in the $[-1, 1]$ domain. This metric allows measuring and comparing the diversity of scientific knowledge of the countries analyzed and demonstrating that, in general, all those countries defined as Developed by the World Bank has a metric value greater than zero, while for the rest of the countries that value is less than zero (negative). In addition, a strong correlation was found between this indicator and the GDP per capita of each of the countries.

Once the most important disciplines were defined, an algorithm was designed to predict the development of new scientific disciplines at the institutional level, through a transformation of the disciplinary complement networks, known as Cartography of Knowledge. The results shown that the position (Centrality) of the undeveloped disciplines is more important than the Activity of their neighboring developed disciplines. The prediction of the algorithm was quantified obtaining an average sensitivity of 0,84. In addition, a web tool for the analysis of co-authorship and disciplinary complement networks was developed, which allows classifying authors according to their importance and visualizing the relationships between them.

Finally, we have extended the study to an online social network, measuring political polarization through an algorithm based on the position of users in Retweet networks for the second round of presidential elections Chilean 2017. Based on a few users whose political opinion is known, we have estimated the opinion of the rest of the users participating in the conversation. With this it was observed that the day before the presidential election was when the system reaches its maximum polarization value and the opinions of the users of that day correlate well with the final results of the electoral elections.

Índice

Agradecimientos	VII
Resumen	IX
Abstract	XI
1. Introducción	1
I Marco teórico	5
2. Sistemas Complejos	7
2.1. Introducción a los sistemas complejos	7
2.2. Propiedades de los sistemas complejos	8
2.2.1. Emergencia	8
2.2.2. Interdependencia	9
2.2.3. Auto-organización	9
2.2.4. Adaptatividad	9
2.3. Introducción a las redes complejas	10
2.3.1. Redes Pesadas	11
2.3.2. Redes Dirigidas	11
2.3.3. Árbol	11
2.3.4. Caminos y Componentes	12
2.3.5. Red bipartita	12
2.3.6. Redes de mundo pequeño	13
2.3.7. Medidas de centralidad	14
II Metodologías	21
3. Redes Sociales	23
3.1. Twitter	24
3.2. Colaboración Científica	25

3.3. Actividad, Respuesta del Sistema, Eficiencia y Centralidad . . .	26
3.4. Características de las redes estudiadas	27
III Resultados	29
4. Redes sociales presenciales: Redes de coautoría	31
4.1. Introducción	33
4.2. Metodología	34
4.3. Actividad, Respuesta del Sistema y Centralidad en las redes de coautoría	35
4.4. Discusión	41
5. Redes de complemento disciplinar	43
5.1. Introducción	43
5.2. Metodología para diferenciar estructuras disciplinares entre países	45
5.3. Construcción del conocimiento científico	47
5.4. Metodología del índice de complejidad disciplinar	47
5.5. Resultados	54
5.6. Discusión	57
6. Modelo predictivo de disciplinas científicas basado en redes de complemento disciplinar.	59
6.1. Introducción	59
6.2. Descripción del modelo	60
6.3. Elección de universidades y set de datos	66
7. Herramienta de visualización de redes de colaboración cien- tífica	73
7.1. Método de construcción	74
7.2. Redes de Personas	74
7.3. Redes de Instituciones	75
7.4. Redes de Disciplinas	76
7.5. Red Conecta	77
7.6. Ranking de Autores	79
7.7. Datos de un autor	81
8. Redes sociales virtuales: Polarización en Twitter	85
8.1. Introducción	85
8.2. Sistema Electoral	87
8.2.1. Sistema electoral chileno	87
8.2.2. Descripción de los datos	89

8.3. Metodología	89
8.3.1. Redes de retweet	89
8.3.2. Inferiendo la opinión de los usuarios	92
8.3.3. Usuarios de Élite	93
8.4. Resultados	95
8.5. Discusión	101
9. Conclusiones	103
9.1. Trabajo a futuro	105
Bibliografía	107
IV Apéndices	113
A. Aprendizaje automático	115
A.1. Introducción al aprendizaje automático	115
A.2. Aprendizaje Supervisado	116
A.2.1. Máquinas de Soporte Vectorial	117
A.3. Aprendizaje No Supervisado	118
A.3.1. Análisis de Componentes Principales	119
A.3.2. K-Medios	120
B. Visualización de redes de complemento disciplinar	121
B.1. Disciplinas correspondientes a cada nodo	121
B.2. Argentina	136
B.3. Brasil	138
B.4. Chile	140
B.5. Colombia	142
B.6. Cuba	144
B.7. Finlandia	146
B.8. Grecia	148
B.9. Islandia	150
B.10. Israel	152
B.11. Nueva Zelandia	154
B.12. Polonia	156
B.13. Portugal	158
B.14. España	160
B.15. Venezuela	162
C. Tabla de conversión disciplinar	165

D. Ponderación de importancia disciplinar por país	179
--	-----

Índice de figuras

2.1. Características de los sistemas complejos (Clemens, 2019). . .	8
2.2. Representación gráfica de un grafo con autoenlaces.	10
2.3. Ejemplo de una red dirigida.	11
2.4. Representación gráfica de un árbol.	12
2.5. Ejemplo de una red con dos componentes separados.	13
2.6. Ejemplo de una red bipartita. A la izquierda se encuentra la partición U y a la derecha la partición V	14
2.7. Ejemplo de grado para el nodo central con $k = 3$	15
2.8. Ejemplo de un nodo central con grado de entrada $k_{in} = 1$ y grado de salida $k_{out} = 2$	15
2.9. Ejemplo de <i>Pagerank</i> en una red con tres nodos.	16
2.10. Ejemplo de una grafo de siete nodos para calcular el coeficiente de cercanía.	18
2.11. Ejemplo de una grafo de seis nodos para calcular el coeficiente de intermediación.	19
3.1. Estructura de una publicación científica indexada en la Web of Science de Thomson Reuters.	26
4.1. Estructura de metadatos de una publicación indexada en WoS para la construcción de redes de coautoría.	32
4.2. Encabezado de un artículo científico, firmado por G. Olivares, J.P. Cárdenas, J.C. Losada y J. Borondo.	32
4.3. Ejemplo de una red de co-autoría para tres autores chilenos en el período 2008-2018, cuyos nodos son presentados en color rojo. Los nodos representan los autores, su tamaño el número de publicaciones realizadas y el color la institución principal a la cual está afiliado. Los enlaces indican la co-autoría entre dos autores y el tamaño del enlace el número de publicaciones en las que han sido co-autores.	33
4.4. Distribución de eficiencia de los autores en escala logarítmica.	36

4.5. Diagrama de Dispersión y correlación entre Pagerank y número de publicaciones de los autores en la red, correspondientes a cada una de las seis áreas de la OECD.	37
4.6. Diagrama de Dispersión y correlación entre Grado y número de publicaciones de los autores en la red, correspondientes a cada una de las seis áreas de la OECD.	38
4.7. Diagrama de Dispersión y correlación entre Pagerank y número de publicaciones (izquierda) y entre Grado y número de publicaciones (derecha) para el área <i>Natural Sciences</i> eliminando los 11 autores con más de 200 publicaciones.	39
4.8. Diagrama de Dispersión y correlación entre Pagerank y número de citas de los autores en la red, correspondientes a cada una de las seis áreas de la OECD.	40
4.9. Diagrama de Dispersión y correlación entre Pagerank y Eficiencia de los autores en la red, correspondientes a cada una de las seis áreas de la OECD.	41
5.1. Diagrama de datos de una publicación científica. Se debe considerar que esta parte sólo incluye la parte de las revistas y áreas científicas.	44
5.2. Ejemplo de generación de una red de complemento disciplinar para tres publicaciones distintas.	44
5.3. Ejemplo de red de complemento disciplinar generada para las publicaciones de tres autores chilenos. Cada nodo representa una disciplina científica, el color representa un Área de la OECD y el tamaño el número de publicaciones desarrollada en cada disciplina. Los enlaces representan los vínculos entre dos disciplinas científicas y el tamaño del enlace el número de veces que ambas disciplinas interaccionaron.	45
5.4. ICD por país. En rojo se muestran los países en vías de desarrollo, en morado, países que serán desarrollados en el futuro y en azul, países desarrollados de acuerdo a la clasificación del Banco Mundial del 2019.	55
5.5. Gráfico de dispersión entre el PIB per cápita para cada país en el año 2018 y su ICD respectivo. Cada punto representa un país.	56
5.6. Gráfico de dispersión entre el PIB per cápita efectivo en I+D para cada país del estudio en el año 2018 y su ICD respectivo. Cada punto representa un país.	57
6.1. Red disciplinar transformada a Cartografía del Conocimiento en Chile.	61

6.2. Ejemplo de las disciplinas desarrolladas (de color distinto a gris) y con potencial de desarrollo (color gris) para la Pontificia Universidad Católica de Valparaíso en el período 2008-2014. El tamaño del nodo es proporcional al número de publicaciones por disciplina.	62
6.3. Ejemplo de las disciplinas de alto potencial de desarrollo para la Pontificia Universidad Católica de Valparaíso para el año 2015, utilizando el período 2008-2014 con el parámetro $\alpha = 2$	63
6.4. Esquema de proyección de áreas desarrolladas por una universidad (en azul) sobre la CCC.	64
6.5. Evolución del promedio de la exactitud (izquierda) y la sensibilidad (derecha) de las 20 universidades seleccionadas de acuerdo a la evolución del parámetro α	69
6.6. Barra de errores de los promedios de todas las universidades para cada valor de α y del modelo nulo.	70
7.1. Interfaz de búsqueda de RedSearch.	73
7.2. Esquema de datos y listado de autores para la Publicación 1.	74
7.3. Red de autores para Eric Goles, Juan Pablo Cárdenas y Gastón Olivares. El color del nodo es de acuerdo a la afiliación asociada, su tamaño es proporcional al número de publicaciones y el peso del enlace corresponde al número de interacciones entre dos autores.	75
7.4. Esquema de datos para una red de instituciones.	76
7.5. Red de Instituciones de acuerdo a la afiliación de los autores que surgen a partir de la búsqueda de Eric Goles, Juan Pablo Cárdenas y Gastón Olivares.	77
7.6. Esquema de datos para red de disciplinas.	78
7.7. Red de Disciplinas de acuerdo a las disciplinas asociadas a las publicaciones de Eric Goles, Juan Pablo Cárdenas y Gastón Olivares.	79
7.8. Estructura básica de un grafo Conecta. En la figura se observan dos autores y una institución, en donde ambos autores están conectados entre sí debido a una coautoría en conjunto y además por estar afiliados en la misma institución.	80
7.10. Ranking de autores para la red de la disciplina Biología en el año 2017 para Chile.	80
7.9. Ejemplo de camino mas corto para los autores Eric Goles y Gastón Olivares.	82
7.11. Resumen del autor Leonardo Bacigalupe Cirillo.	83
7.12. Preferencia de conexión del autor Leonardo Bacigalupe Cirillo.	83
7.13. Comparativa de las disciplinas desarrolladas por Leonardo Bacigalupe Cirillo y sus coautores.	84

8.1. Resultados de la primera ronda de las elecciones presidenciales chilenas del 2017.	88
8.2. Resultados de la segunda ronda de las elecciones presidenciales del 2017.	89
8.3. Redes de retweets de la segunda ronda de las elecciones presidenciales chilenas del 2017, correspondiente a la campaña electoral del período del 11 al 16 de diciembre (a) y del 11 al 17 de diciembre (b). El tamaño de los nodos es proporcional a su grado de entrada k_{in} y sus colores están asociados a la opinión de acuerdo a la escala que se muestra en la leyenda: -1 Guillier, 0 Neutral y 1 Piñera.	91
8.4. Distribución del grado de entrada y de salida correspondiente a la red de <i>reweet</i> para el período del 11 al 17 de diciembre de 2017.	92
8.5. Subred de la red de <i>retweets</i> correspondiente a la conversación de Twitter de las elecciones chilenas del período 11 al 17 de diciembre de 2017, correspondientes a los 160 nodos con $k_{in} > 200$. Los nodos de la élite asociados a Piñera en rojo y los asociados a Guillier en azul. El resto de los nodos en gris no están incluidas en las élites. El tamaño de los nodos corresponde a su respectivo k_{in}	94
8.6. Convergencia de las opiniones inferidas a través de un proceso iterativo correspondiente al modelo de formación de opinión para 100 nodos seleccionados aleatoriamente. Se debe notar que muchos nodos obtienen las mismas dinámicas de convergencia.	95
8.7. Paneles superiores: distribución de opinión estimada para las redes de <i>retweets</i> correspondiente a la campaña electoral del período 11 al 16 de diciembre del 2017 (izquierda) y a la campaña electoral del período 11 al 17 de diciembre 2017 (derecha). Panel inferior: distribución de opinión de las redes de <i>retweets</i> correspondiente a la campaña electoral del período 11 al 17 de diciembre de 2017 eliminando a aquellos usuarios que sólo participaron durante el día de la votación.	97
8.8. Evolución de la distribución de probabilidad de las opiniones para los valores X: -1, +1, >0 y <0 (arriba) y los índices de polarización política, ρ , para las redes de <i>retweet</i> comenzando el 11 de diciembre hasta el día D de las elecciones presidenciales chilenas de 2017. Los valores de la línea punteada corresponde a los de la campaña electoral del período 11 al 17 de diciembre de 2017 sin tomar a aquellos usuarios que sólo participaron el día de la votación.	98

8.9. Evolución en el porcentaje de menciones (arriba) y el índice de heterogeneidad (abajo), para las redes de <i>retweet</i> comenzando el 11 de diciembre hasta el día D de las elecciones presidenciales chilenas del 2017. En el panel superior, la línea punteada corresponde a la campaña electoral del período 11 al 17 de diciembre de 2017 sin considerar aquellas cuentas de usuarios que sólo participaron el día de la votación.	100
A.1. Diagrama de flujo de un algoritmo de aprendizaje automático supervisado.	116
A.2. Funcionamiento de una máquina de soporte vectorial (Meyer y Wien, 2001).	118
A.3. Visualización de los datos antes y después de aplicar PCA (Harrington, 2012)	119
A.4. Diagrama de flujo del algoritmo k-medios.	120
B.1. Identificador para cada una de las disciplinas dentro de la visualización.	122
B.2. Identificador para cada una de las sub disciplinas dentro de la visualización.	133
B.3. Red de complemento disciplinar para Argentina en el nivel disciplinar.	136
B.4. Red de complemento disciplinar para Argentina en el nivel sub disciplinar.	136
B.5. Red de complemento disciplinar para Argentina en el nivel de áreas de la OECD.	137
B.6. Red de complemento disciplinar para Brasil en el nivel disciplinar.	138
B.7. Red de complemento disciplinar para Brasil en el nivel sub disciplinar.	138
B.8. Red de complemento disciplinar para Brasil en el nivel de áreas de la OECD.	139
B.9. Red de complemento disciplinar para Chile en el nivel disciplinar.	140
B.10. Red de complemento disciplinar para Chile en el nivel sub disciplinar.	140
B.11. Red de complemento disciplinar para Chile en el nivel de áreas de la OECD.	141
B.12. Red de complemento disciplinar para Colombia en el nivel disciplinar.	142
B.13. Red de complemento disciplinar para Colombia en el nivel sub disciplinar.	142

B.14.Red de complemento disciplinar para Colombia en el nivel de áreas de la OECD.	143
B.15.Red de complemento disciplinar para Colombia en el nivel disciplinar.	144
B.16.Red de complemento disciplinar para Colombia en el nivel sub disciplinar.	144
B.17.Red de complemento disciplinar para Colombia en el nivel de áreas de la OECD.	145
B.18.Red de complemento disciplinar para Finlandia en el nivel disciplinar.	146
B.19.Red de complemento disciplinar para Finlandia en el nivel sub disciplinar.	146
B.20.Red de complemento disciplinar para Finlandia en el nivel de áreas de la OECD.	147
B.21.Red de complemento disciplinar para Grecia en el nivel disciplinar.	148
B.22.Red de complemento disciplinar para Grecia en el nivel sub disciplinar.	148
B.23.Red de complemento disciplinar para Grecia en el nivel de áreas de la OECD.	149
B.24.Red de complemento disciplinar para Islandia en el nivel disciplinar.	150
B.25.Red de complemento disciplinar para Islandia en el nivel sub disciplinar.	150
B.26.Red de complemento disciplinar para Islandia en el nivel de áreas de la OECD.	151
B.27.Red de complemento disciplinar para Israel en el nivel disciplinar.	152
B.28.Red de complemento disciplinar para Israel en el nivel sub disciplinar.	152
B.29.Red de complemento disciplinar para Israel en el nivel de áreas de la OECD.	153
B.30.Red de complemento disciplinar para Nueva Zelandia en el nivel disciplinar.	154
B.31.Red de complemento disciplinar para Nueva Zelandia en el nivel sub disciplinar.	154
B.32.Red de complemento disciplinar para Nueva Zelandia en el nivel de áreas de la OECD.	155
B.33.Red de complemento disciplinar para Polonia en el nivel disciplinar.	156
B.34.Red de complemento disciplinar para Polonia en el nivel sub disciplinar.	156

B.35.Red de complemento disciplinar para Polonia en el nivel de áreas de la OECD.	157
B.36.Red de complemento disciplinar para Portugal en el nivel disciplinar.	158
B.37.Red de complemento disciplinar para Portugal en el nivel sub disciplinar.	158
B.38.Red de complemento disciplinar para Portugal en el nivel de áreas de la OECD.	159
B.39.Red de complemento disciplinar para España en el nivel disciplinar.	160
B.40.Red de complemento disciplinar para España en el nivel sub disciplinar.	160
B.41.Red de complemento disciplinar para España en el nivel de áreas de la OECD.	161
B.42.Red de complemento disciplinar para Venezuela en el nivel disciplinar.	162
B.43.Red de complemento disciplinar para Venezuela en el nivel sub disciplinar.	162
B.44.Red de complemento disciplinar para Venezuela en el nivel de áreas de la OECD.	163

Índice de Tablas

2.1.	Camino más corto para cada nodo de la figura 2.10.	19
2.2.	Cálculo de coeficiente de intermediación para el nodo C 2.11 de la red mostrada en la figura 2.10.	20
3.1.	Diferencia de las distintas versiones de la API de Twitter. La versión estándar es gratuita.	25
3.2.	Diferencias entre redes sociales virtuales (elecciones presiden- ciales de Chile del 2017 y las elecciones presidenciales de Bra- sil del 2018) y redes sociales presenciales para (co-autoría de Chile 2017 y co-autoría de Brasil 2018).	28
4.1.	Total de autores chilenos con al menos una colaboración cien- tífica para cada una de las redes de co-autoría de cada área de la OECD. A la derecha se presenta el porcentaje de autores chilenos que pertenece a la componente gigante.	35
5.1.	Número de publicaciones para cada uno de los países estudia- dos en el período 2008-2018.	46
5.2.	FID para cada una de las disciplinas estudiadas.	53
5.3.	ICD para cada uno de los países del estudio.	54
5.4.	PIB per cápita, Porcentaje del PIB per cápita destinado a I+D y PIB per cápita destinado a I+D para cada país.	56
6.1.	Valores T_i , a_i y b_i para cada una de las disciplinas de la figura 6.4.	65
6.2.	Distribución de publicaciones para cada una de las univer- sidades seleccionadas para la evaluación de la metodología, separados en conjunto de datos de disciplinas desarrolladas por una universidad (2008-2014) y conjunto de datos de eva- luación (2015-2018)	67
6.3.	Tabla de contingencia.	67

6.4. Distribución de disciplinas desarrolladas en el período 2008-2014 y 2015-2018, en conjunto con la matriz de confusión y los valores de pruebas diagnósticas exactitud y sensibilidad para cada una de las 20 universidades seleccionadas.	72
A.1. Ejemplo de conjunto de datos para ser utilizados en aprendizaje automático	115
B.1. Asignación del identificador junto a su disciplina científica asociada de la figura B.1	132
B.2. Asignación del identificador junto a su sub disciplina científica asociada de la figura B.2	135
C.1. Tabla de conversión disciplinar	178
D.1. Tabla de ponderación de importancia disciplinar para Argentina, Brasil, Chile y Colombia.	188
D.2. Tabla de ponderación de importancia disciplinar para Cuba, España, Finlandia y Grecia.	196
D.3. Tabla de ponderación de importancia disciplinar para Islandia, Israel, Nueva Zelandia y Polonia.	205
D.4. Tabla de ponderación de importancia disciplinar para Portugal y Venezuela.	214

Capítulo 1

Introducción

Hoy en día el concepto de red social está presente en el imaginario colectivo de una gran parte de la población mundial. Esto gracias a que el desarrollo de la Internet permitió que las conexiones interpersonales fuesen mucho más sencillas para todos, dejando de lado muchas veces la dimensión geográfica.

Fueron los científicos quienes fortalecieron el concepto, no solo por la creación de la Internet con fines inicialmente fueron para defensa (ARPA), pero que después se convirtió en la red de redes que conocemos hoy, la cual se utiliza para la resolución de problemas cotidianos como el consumo de café, un gran aliado que les permitía continuar con sus investigaciones de forma mucho más amena. El problema del café radicaba en que su sala se encontraba en un punto alejado de la mayoría de los laboratorios de investigación y cada científico debía gastar varios minutos de su tiempo en recorrer el edificio para llevarse muchas veces la ingrata sorpresa que su colega de otro laboratorio ya había acabado con todo el café disponible. Esto provocó diversos quiebres a pequeña escala, donde los conflictos y la aparición de grupos con pensamiento distinto (polarizaciones) comenzaron a aparecer. A raíz de este problema, la comunidad de científicos decidió colaborar entre sí y utilizando sus arcaicos ordenadores conectados en red para colocar una videocámara en la sala del café y que todos los miembros del establecimiento tuvieran acceso a esta vista y a una improvisada sala de chat para organizar mucho mejor sus tiempos para beber café (Stafford-Fraser, 2001). Este problema con el café es un simple ejemplo de la importancia que ha cobrado la tecnología como medio para conectar a las personas.

Este tipo de redes sociales *online* no serían más que la proyección de las tradicionales relaciones humanas hacia una dimensión digital. De la misma forma que antiguamente la gente se juntaba en las plazas para conversar, hoy en día lo hace gracias a las primeras herramientas de conversación basadas en texto, como IRC, ICQ, MSN Messenger, llegando hoy en día a sistemas mucho más sofisticados como Whatsapp, que no sólo permiten el intercambio

de texto, sino también de mensajes de voz y vídeo en tiempo real.

Adicionalmente, de la misma forma que los mercaderes ofrecían sus productos en plazas o mercados, hoy en día estas redes sociales no solo permiten el intercambio de información sino que también se puede realizar a través de ellas la comercialización de productos y servicios. Plataformas digitales como Ebay o Amazon son ejemplos de lo anterior.

Todo este «nuevo mundo digital» está, a su vez, muy intervenido por la política y por mucho que se piense que el desarrollo científico y tecnológico debe ser neutral al respecto, es inevitable observar que esto está lejos de ser así. Hoy en día, Ciencia, Tecnología, Política y Economía se entrelazan de una forma incluso más rica que antes.

Hasta hace pocas décadas el conocimiento técnico era para unos pocos, debido a lo complicado que era su divulgación y a que un porcentaje muy bajo de la población estaba alfabetizada. Hoy en día eso también ha cambiado, gracias no solo a la baja tasa de analfabetismo, sino debido al surgimiento de la nueva era digital, la era de la Información, que ha hecho posible compartir estos conocimientos de una forma mucho más sencilla, de manera tal que prácticamente cualquier humano en cualquier sitio del planeta pueda acceder a ella desde prácticamente cualquier máquina con conexión a internet.

Es en este contexto de las tecnologías de la Información y en el de una economía basada en la comercialización de artefactos de información, donde se enmarca esta tesis, en donde partimos de la hipótesis de que los sistemas sociales pueden modelizarse como redes o grafos cuyos componentes poseen características propias cuantificables como son su actividad en el sistema u otras referidas a su posición en éste, tales como su centralidad. Por otro lado, a partir de las relaciones entre dichos componentes, emergen en éste ciertas respuestas del sistema hacia ellos. A partir de estas características, tanto locales como globales, es posible construir algoritmos e índices que permiten analizar y comparar estas redes sociales a diferentes niveles, desde individuos a organizaciones que los agrupan como instituciones e incluso países.

Dentro de la naturaleza de estas relaciones se pueden definir dos tipos, las positivas, las cuales dentro del desarrollo de esta tesis se expresan como la colaboración o la influencia positiva de un usuario a otro; y las negativas, las que en el contexto de redes sociales se puede representar a través de la polarización política.

El objetivo de este trabajo es proponer distintas metodologías para caracterizar diversos tipos de sistemas sociales a través de medidas de Actividad, Respuesta del Sistema y Centralidad que serán establecidas dependiendo de la estructura de las redes estudiadas. Además, se espera que este tipo de medidas establezcan un vínculo entre dos fenómenos antagónicos, aunque estrechamente relacionados, como son la colaboración y la polarización. El primer fenómeno describe la forma en que un grupo de individuos trabaja en conjunto en busca de un objetivo en común. Por otro lado, el fenómeno

de la Polarización rompe esta sinergia global provocando la separación del colectivo en dos o más grupos antagónicos, aunque en su interior existe la colaboración. A pesar de esta relación, ambos fenómenos, como es de esperar, poseen propiedades diferentes.

Para abordar el estudio de ambos fenómenos, en sus diferencias y similitudes, se analizará la colaboración científica y la polarización política como ejemplos de estos fenómenos.

Para estudiar la colaboración científica se utilizará la base de datos de Thomson Reuters, una de las más completas respecto al almacenamiento de toda la información científica disponible. En este caso, este trabajo asume que la colaboración científica ocurre tanto en un nivel personal (donde los investigadores se conocen y trabajan juntos), donde además es posible generar redes a distintos niveles, incluyen niveles institucionales como de países y redes funcionales, como son el caso de las redes de complemento disciplinar, que son estructuras de red en donde las disciplinas son los nodos y los enlaces se generan a partir de sus vínculos en una publicación científica. Se debe considerar además que a lo largo de este trabajo se generan redes no dirigidas debido a que suponemos que la colaboración es bidireccional.

Para el caso de estudio de la Polarización Política, se utilizarán datos de una red social enteramente digital como es Twitter, la cual en los últimos años se ha visto envuelta en grandes temas de polarización política, como por ejemplo el caso del Brexit, cuyo tema ha sido ampliamente hablado en esta red social y entrego mensajes positivos y negativos respecto a esta política (RCN-Radio, 2019). Es importante destacar que la API de Twitter permite la descarga de *Tweets* de forma gratuita, aunque con ciertas limitaciones. Dentro de estas limitaciones se tiene que para temas específicos se necesita ir descargando los datos en tiempo real, ya que no se permite la descargas de *tweets* antiguos, a menos que se pague por la versión completa.

Esta tesis propone distintas metodologías para la visualización e interpretación de las relaciones sociales mediante técnicas de redes complejas que aparecen en estos dos dominios (científico y político) con el fin de observar ciertos fenómenos sociales comunes a ambas redes así como ciertas particularidades.

La estructura de esta tesis consta de cinco partes: Marco Teórico, Metodologías, Resultados, Conclusiones y Apéndices.

En Marco Teórico se profundizará en la teoría y métodos de las Redes Complejas, partiendo desde los conceptos básicos hasta ciertas propiedades y métricas más sofisticadas utilizado a lo largo de este trabajo.

Por el lado de las Metodologías se describirán los tipos de redes sociales, junto con la definición de los conceptos de Actividad, Respuesta del Sistema y Centralidad utilizados a lo largo de esta tesis

En la sección de Resultados se presentarán los resultados obtenidos para los diversos tipos de sistemas sociales, partiendo por las redes de co-autoría,

en donde se correlacionan las medidas de Actividad y Centralidad para elegir a las medidas más importantes en el desarrollo de los siguientes capítulos.

Además, se propondrá una metodología para cartografiar el conocimiento científico en catorce países, donde, a través de redes de complemento disciplinar, se estudiarán las interacciones que aparecen entre diversas disciplinas científicas y posteriormente definir un ranking disciplinar. Además se definirá una métrica que permita diferenciar que tan conectado está el desarrollo del conocimiento científico de cada país con el ranking de disciplinas propuesto.

En el siguiente capítulo se entrega un modelo predictivo de disciplinas científicas basado en redes de complemento disciplinar, cuyo estudio fue aplicado en veinte universidades chilenas en el período 2008-2018.

En el capítulo de redes sociales *online* se estudiará la polarización política en Twitter, en particular en la aplicación de un algoritmo de difusión de información a través de redes de *Retweets* para las elecciones presidenciales de Chile el 2018 y así estimar la opinión política de los usuarios.

Finalmente, se enumeran las conclusiones más relevantes de la tesis y posteriormente se incluyen las referencias y diversos apéndices con información relevante para la construcción de cada capítulo.

Parte I

Marco teórico

La primera parte de esta tesis consiste en una descripción general de los conceptos básicos más utilizados a lo largo de los siguientes capítulos, en donde se abordarán temas de Sistemas Complejos y Redes Complejas.

Capítulo 2

Sistemas Complejos

2.1. Introducción a los sistemas complejos

Cada vez que interactuamos con nuestro entorno podemos observar que a cada instante suceden diversos acontecimientos, algunos predecibles y otros totalmente fortuitos. Cuando vemos un juego de billar apreciamos que un jugador es capaz de golpear la bola con el taco con una fuerza y ángulo que permite que la bola haga exactamente lo que el jugador quiere. Este tipo de sistemas son muy predecibles debido a que existe un proceso previo de aprendizaje en donde el jugador sabe que tiene que utilizar cierta fuerza sobre cierto ángulo para lograr su objetivo. Por lo mismo, es posible modelar este tipo de sistemas con ecuaciones simples, como es la ecuación del tiro parabólico. Por otro lado, cuando queremos ser millonarios y participamos de las tradicionales loterías, elegimos un conjunto de números y probamos suerte para ver si ganamos. Este tipo de sistema es completamente aleatorio e imposible de predecir, tal como puede ser la distribución de partículas de oxígeno en una habitación. Tradicionalmente los ejemplos descritos anteriormente podemos clasificarlos como sistemas simples y sistemas aleatorios, sin embargo, en la intersección de estos dos mundos es posible hallar otro tipo de sistemas, como lo son los Sistemas Complejos.

La RAE define como Sistema a un conjunto de reglas o principios sobre una materia racionalmente enlazados entre sí y Complejo a algo que se compone de diversos elementos. Por lo tanto podemos inferir que un Sistema Complejo es un sistema compuesto de muchos elementos que pueden interactuar entre sí, tales como la infraestructuras de redes eléctricas, sistemas de transportes o incluso el cerebro de los seres vivos. Sin embargo existen diversas definiciones y detallaremos algunas de ellas.

Hebert Simon (Simon, 1977) define como sistema complejo a aquellos compuestos por un gran número de partes que interactúan en modo no simple, en donde la multiplicidad es el principal atributo que describe a la complejidad.

Por otro lado, Bar-Yam, Yaneer (Bar-Yam, 2002) considera a los sistemas complejos como un enfoque de la ciencia que estudia cómo las relaciones entre las partes dan origen a los comportamientos colectivos de un sistema, y cómo el sistema interactúa y forma las relaciones con su entorno. Además, Murray (Murray, 1995) propone que los sistemas complejos se ubican en un punto intermedio entre el orden y el desorden.

Sin embargo, existen una serie de características que un sistema debe poseer para ser complejo y estas son descritas por Bar-Yam (Bar-Yam, 2002) y Torres (Torres, 2010), entre las que destacamos los conceptos de emergencia, interdependencia, auto-organización y adaptatividad (Ver figura 2.1), los que se describirán a continuación.

Figura 2.1: Características de los sistemas complejos (Clemens, 2019).

2.2. Propiedades de los sistemas complejos

2.2.1. Emergencia

El primer concepto introducido corresponde al de emergencia, el que describe la relación entre los objetos de menor nivel (detalles) hasta los de mayor nivel (vista del todo) (Bar-Yam, 2002). Se puede ejemplificar mejor este concepto si hacemos referencia a la polarización de los individuos, entendiendo

por polarización a la propiedad social en donde dos grupos de individuos tienen una opinión totalmente opuesta sobre un tema en específico. El concepto de polarización no es algo que pueda asignarse de forma individual sin considerar al resto de las partes, es efectivamente cuando emerge el concepto de sociedad a causa de sus interacciones y es ahí cuando recién se puede hablar de polarización como consecuencia de la suma de todas las partes de menor nivel.

2.2.2. Interdependencia

El concepto de interdependencia habla de cómo las interacciones de cada una de las partes del sistema dan origen, a través de las relaciones, a comportamientos colectivos (Bar-Yam, 2002). Por ejemplo, un trabajador de una fábrica de sillas puede dedicarse a cortar madera. Esto por sí sólo no significa nada, sin embargo en otra línea de producción tenemos a alguien que une las piezas y finalmente otra persona que la barniza, dando como resultado una silla. Claramente si uno de los trabajadores se niega a realizar su trabajo, el producto final no podría ser terminado, por lo que es esta interdependencia la cual puede afectar negativamente al resultado final.

2.2.3. Auto-organización

Se entiende por auto-organización al proceso mediante el cual un sistema natural se reproduce a sí mismo mediante su propia lógica y los elementos que lo componen (Bar-Yam, 2002). Se puede ejemplificar este concepto cuando observamos el movimiento de una bandada de aves, en donde cada uno de ellos se organiza sin necesidad que exista una inteligencia superior, cuyos movimientos son realizados de acuerdo a lo que hace su vecino más cercano y es posible ver, de forma macro, un organismo totalmente auto organizado con movimientos particulares.

2.2.4. Adaptatividad

Este concepto va muy asociado al de la auto-organización, ya que indica que un sistema complejo debe tener la capacidad de cambiar y aprender de la experiencia (Waldrop, 1992).

Lo que describimos corresponden a los conceptos claves para entender cómo está estructurado un Sistema Complejo, sin embargo, a lo largo de esta tesis, trabajaremos con Redes Complejas, que consisten en grafos o redes que son una abstracción de los sistemas complejos, en los cuales los vértices son los componentes del sistema y las aristas representan sus relaciones, los cuales serán presentados en la siguiente sección.

2.3. Introducción a las redes complejas

Se puede describir una grafo o una red como una colección de puntos (vértices) unidos a otros por aristas o arcos para representar relaciones binarias entre elementos de un conjunto (Newman, 2010), un ejemplo se puede observar en la Figura 2.2.

Una de las formas de representar una red es a través de una matriz de ad-

Figura 2.2: Representación gráfica de un grafo con autoenlaces.

yacencia. Una matriz de adyacencia se define como una matriz cuadrada que se utiliza para representar relaciones binarias. Si consideramos una red sin dirigir con n vértices, en donde la etiqueta de estos vértices corresponden a números naturales sucesivos, tenemos la siguiente ecuación 2.1.

$$A_{ji} = \begin{cases} 1 & \text{si existe un enlace entre } i \text{ y } j \\ 0 & \text{en otro caso} \end{cases} \quad (2.1)$$

Es posible representar múltiples enlaces y auto enlaces utilizando esta matriz de adyacencia. Un multienlace es representado por una configuración correspondiente a un elemento A_{ij} multiplicado por el número de enlaces N . De esta forma existirán tantos enlaces entre i y j de acuerdo al número N dado. Para el caso de los autoenlaces, estos son representados cuando $i = j$ con $A_{ij} = A_{ji} = 1$, debido a que se están representando dos enlaces en direcciones opuestas. En la siguiente imagen se observa la representación de la red de la ecuación 2.2 en una matriz de adyacencia.

$$A = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \quad (2.2)$$

2.3.1. Redes Pesadas

A pesar de que muchas de las redes que se estudian tienen conexiones simples (hay o no conexión), existen otra gran cantidad en donde es muy útil representar los enlaces con peso. Por ejemplo, estos enlaces pueden representar la cantidad de mensajes enviados entre dos personas. Estas conexiones pueden ser representadas en los elementos de la matriz de adyacencia, en donde el valor A_{ij} corresponde al peso del enlace.

2.3.2. Redes Dirigidas

Las redes dirigidas son aquellas en donde el enlace tiene dirección, como en el caso del ejemplo de la Figura 2.3.

Figura 2.3: Ejemplo de una red dirigida.

Una de las propiedades de las redes dirigidas es que generalmente su matriz de adyacencia suele ser asimétrica. Para el caso del ejemplo anterior, la matriz de adyacencia se representaría como en la ecuación 2.3.

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 1 \end{bmatrix} \quad (2.3)$$

2.3.3. Árbol

Un árbol es una red conectada y no dirigida que no contiene bucles cerrados. Cuando hablamos de red conectada nos referimos a que cada vértice de la red puede alcanzar a cualquier otro por un camino entre todos los vértices

de la red. En la figura 2.4 se puede observar en color negro la representación visual de un árbol, mientras que el enlace adicional en gris crea un bucle, por lo que si se adiciona ese enlace la red deja de ser un árbol.

Figura 2.4: Representación gráfica de un árbol.

2.3.4. Caminos y Componentes

Definiremos como camino en una red como la secuencia de vértices, en donde cada par de vértices consecutivos de una secuencia está conectado por un enlace de la red. Puede que no hayan caminos entre todos los pares de vértices dentro de la red. Cuando esto sucede, los vértices son divididos en subgrupos que no tienen conexiones entre ellos, tal como se puede observar en la figura 2.5. Los subgrupos de la figura anterior son llamados componentes, el cual se puede definir como un subconjunto de vértices de la red que tiene al menos un vértice asociado con otro del mismo subconjunto de la red. Dentro de las medidas que se utilizan en los caminos tenemos el diámetro de la red, que es la longitud del camino más corto para unir los dos vértices más alejados de la red. En el caso de que la red no esté conectada, el diámetro es el máximo de todos los diámetros para su componente conectada. También se puede definir su diámetro a través de la excentricidad máxima de los nodos, en donde la excentricidad de un nodo es el largo máximo de un camino corto entre él y otro nodo de la red. En el caso de que el nodo se encuentre aislado, su excentricidad es cero (Assenov et al., 2008).

2.3.5. Red bipartita

Las redes o grafos bipartitos son grafos donde sus nodos o vértices pueden ser separados por dos particiones, U y V , que contiene enlaces E entre los

Figura 2.5: Ejemplo de una red con dos componentes separados.

nodos de una partición con los de la otra (Aksoy et al., 2016). Se define como $G = (U, V, E)$ con $U \cap V = \emptyset$ y $E \subseteq U \otimes V$.

Un ejemplo de grafo bipartito se muestra en la figura 2.6.

Para el caso de un ejemplo de tipo de red, se puede considerar el grafo de coautorías, en el que los nodos de la partición U serían representados por los autores, mientras que los nodos de la partición V serían las disciplinas científicas de la revista en la cual se haya publicado.

2.3.6. Redes de mundo pequeño

Las redes de mundo pequeño son un tipo de grafo en donde la mayoría de los nodos no están enlazados entre ellos, sin embargo la mayoría ellos pueden ser alcanzados desde cualquier punto por un número pequeño de pasos (Zaidi, 2012). Este tipo de grafos se pueden encontrar en diversos sistemas reales, como las redes sociales (Wasserman y Faust, 1994), sistemas de transporte (Guimera et al., 2005) y muchas más. Este tipo de redes tienen dos propiedades estructurales, el efecto del pequeño mundo y el coeficiente de agrupamiento (Watts y Strogatz, 1998). El efecto de mund pequeño tiene relación a un experimento del psicólogo social Stanley Milgran (Milgram, 1967) que llegó a la conclusión que era posible conectar a dos personas de Estados Unidos con una media de seis saltos, a lo que se denominó seis grados de separación.

Por otro lado, estas redes tienen un coeficiente de agrupamiento alto, debido a que tienen la particularidad de agrupar cliques, en donde un clique es un subconjunto de nodos dentro de un grafo dirigido, en donde cada uno de ellos están conectados entre si. Además, muchos de los pares de nodos pueden ser conectados a través de un camino corto.

Figura 2.6: Ejemplo de una red bipartita. A la izquierda se encuentra la partición U y a la derecha la partición V .

2.3.7. Medidas de centralidad

A continuación se definirán las medidas de centralidad más relevantes que se han utilizado a lo largo de esta tesis, entre los que detallaremos el coeficiente de agrupamiento (*Clustering Coefficient*), grado (*Degree*), *Pagerank*, coeficiente de cercanía (*Closeness Centrality*) y coeficiente de intermediación (*Betweenness Centrality*).

2.3.7.1. Grado

Utilizando la definición de (Diestel, 2005), para redes no dirigidas, el grado de un nodo de un nodo n es el número de enlaces totales de n . Finalmente, la distribución de grado de la red entrega el número de nodos con grado k para $k = 0, 1, \dots$. A modo de ejemplo, en la figura 2.7 se puede observar al nodo central con un grado $k=3$, mientras que el resto de los nodos tienen grado $k = 1$ debido a que poseen sólo un enlace cada uno.

Para redes dirigidas, el grado de entrada de un nodo n es el número de enlaces entrantes y el grado de salida el número de enlaces salientes de dicho nodo. Para estos casos existe el homólogo para la distribución de grado de

Figura 2.7: Ejemplo de grado para el nodo central con $k = 3$.

las redes no dirigidas, sin embargo acá se utiliza la distribución de grado de entrada y la distribución de grado de salida. En la figura 2.8 se puede observar que el nodo central tiene dos enlaces de salida y un enlace de entrada, por lo que su grado de salida $k_{out} = 2$, mientras que su grado de entrada $k_{in} = 1$.

Figura 2.8: Ejemplo de un nodo central con grado de entrada $k_{in} = 1$ y grado de salida $k_{out} = 2$.

2.3.7.2. Coeficiente de agrupamiento

Utilizando la definición de (Watts y Strogatz, 1998), el coeficiente de agrupamiento es una medida de grado en donde los nodos de una red tienden a agruparse entre ellos, o visto de otra manera, es la capacidad que tienen tres nodos distintos de formar triángulos.

En redes no dirigidas, el coeficiente de agrupamiento C_n de un nodo n se define como $C_n = \frac{2e_n}{k_n(k_n-1)}$, donde k_n es el número de nodos vecinos de n y e_n es el número de pares conectados entre todos los nodos vecinos de n . En redes dirigidas, la definición cambia a $C_n = \frac{e_n}{k_n(k_n-1)}$.

Para ambos casos, el coeficiente de agrupamiento es el ratio N/M , donde N es el número de enlaces entre los nodos vecinos de n y M es el número máximo de enlaces que pueden existir entre todos los nodos vecinos al nodo n . Debido a lo anterior, este coeficiente siempre se mueve entre 0 y 1. Por otro lado, se puede definir la distribución de coeficiente de agrupamiento promedio como el promedio de los coeficientes de agrupamiento de todos los nodos n con k vecinos y se define de la siguiente manera: 2.4.

$$C = \frac{1}{n} \sum_{i=1}^n C_i \quad (2.4)$$

2.3.7.3. Pagerank

El algoritmo *Pagerank* entrega un ranking de importancia relativa de páginas web, en donde cada sitio web está representada dentro de un grafo de la web (Page et al., 1998). La idea general del algoritmo consiste en que si un sitio A enlaza a un sitio B , significa que el sitio A le entregaría un voto a la página B . Sin embargo la importancia de la página que entrega el voto también pondera una importancia, y su expresión matemática está dada por la ecuación 2.5

$$PR(A) = (1 - d) + d \sum_{i=1}^n \frac{PR(i)}{C(i)} \quad (2.5)$$

donde $PR(A)$ es el *Pagerank* de una página A , d es un factor de amortiguación que fluctúa entre 0 y 1, $PR(i)$ son los valores de *Pagerank* que tienen cada página i enlazada a la página A y $S(i)$ es el número total de enlaces salientes de la página i . A modo de ejemplo se considerará una red de tres nodos mostrados en la figura 2.9, en donde su factor de amortiguación $d = 0,7$ y el número total de nodos $N = 3$.

Figura 2.9: Ejemplo de *Pagerank* en una red con tres nodos.

De esta forma se tiene:

$$PR(A) = (1 - d)(1/N) + d(PR(C)/2) \quad (2.6)$$

$$PR(B) = (1 - d)(1/N) + d(PR(A)/1 + PR(C)/2) \quad (2.7)$$

$$PR(C) = (1 - d)(1/N) + d(PR(B)/1) \quad (2.8)$$

Reemplazando los valores

$$PR(A) = 0,1 + 0,35(PR(C)) \quad (2.9)$$

$$PR(B) = 0,1 + 0,7(PR(A)) + 0,35(PR(C)) \quad (2.10)$$

$$PR(C) = 0,1 + 0,7(PR(B)) \quad (2.11)$$

Y finalmente resolviendo

$$PR(A) = 0,2314 \quad (2.12)$$

$$PR(B) = 0,3933 \quad (2.13)$$

$$PR(C) = 0,3753 \quad (2.14)$$

Las ecuaciones anteriores indican que el sitio B es aquel sitio más importante dentro del grafo, si consideramos el nivel de importancia de acuerdo a su Pagerank mas elevado. Esto también se puede verificar debido a que el sitio B es quien posee la mayor cantidad de enlaces.

2.3.7.4. Coeficiente de cercanía

El coeficiente de cercanía es una medida de que tan rápido la información se propaga de un nodo a otro en la red (Newman, 2003). El coeficiente de cercanía $C_c(n)$ de un nodo n está definido como el recíproco del promedio del camino más corto y se calcula de de acuerdo a la ecuación 2.15.

$$C_c(n) = \frac{1}{avg(L(n, m))} \quad (2.15)$$

donde $L(n, m)$ es el valor de la distancia del camino mas corto entre dos nodos n y m y $avg(L(n, m))$ el promedio de esta distancia. El C_c puede fluctuar entre 0 y 1.

En la figura 2.10 se observa un grafo con siete nodos en total. Si se quisiera calcular el coeficiente de cercanía del nodo C , lo primero que haríamos es calcular el camino más corto desde el nodo c hasta el resto de los nodos, lo cual se puede observar en la tabla 2.1.

Por lo tanto, aplicando la ecuación 2.15, tenemos $avg(L(n, m))$ es igual a $(10)/7 = 1,43$, donde $C_c(C) = 1/1,43$ y finalmente $C_c(C) = 0,6993$.

Figura 2.10: Ejemplo de una grafo de siete nodos para calcular el coeficiente de cercanía.

2.3.7.5. Coeficiente de intermediación

El coeficiente de intermediación de un nodo refleja la cantidad de control que este nodo ejerce sobre las interacciones de otros nodos en la red. Esta medida favorece a los nodos que unen comunidades, en lugar de los nodos que se encuentran dentro de una misma comunidad (Yoon et al., 2006). El coeficiente de intermediación de un nodo n viene definida por la siguiente expresión definida por Brandes (Brandes, 2001).

$$C_b(n) = \sum_{s \neq n \neq t} \frac{\sigma_{st}(n)}{\sigma_{st}} \quad (2.16)$$

donde s y t son nodos de la red distintos a n , σ_{st} es el número de caminos cortos de s a t y $\sigma_{st}(n)$ es el número de caminos cortos desde s a t que pasan por n . Este coeficiente es computado sólo para redes que no contienen enlaces múltiples. El coeficiente de intermediación para cada nodo n está normalizado al dividirlo por el número de nodos pares, excluyendo n , es decir $(M-1)/(M-2)/2$, donde M es el número total de nodos en la componente conectada en donde se encuentra n . Por esta razón, este coeficiente tiene un

[h] Nodo	Camino más corto desde C
A	1
B	1
C	1
D	1
E	2
F	2
G	2

Tabla 2.1: Camino más corto para cada nodo de la figura 2.10.

valor que oscila entre 0 y 1. Para ejemplificar mejor este coeficiente, en la figura 2.11 se muestra un grafo de seis nodos con todos sus nodos conectados en una única componente.

Figura 2.11: Ejemplo de una grafo de seis nodos para calcular el coeficiente de intermediación.

En la tabla 2.2 se muestran los valores para el coeficiente de intermediación del nodo C de la red mostrada en la figura 2.11. Finalmente el coeficiente de intermediación para el nodo C es $C_b(C) = 6$.

2.3.7.6. Coeficiente de asortatividad

La asortatividad es la tendencia observada en redes complejas en donde los nodos más conectados se conectan con otros similares (Bagler y Sinha, 2007). Esta similitud se puede interpretar en términos del grado de los nodos (nodos mas conectados se conectan a otros más conectados), aunque es también se puede utilizar otras propiedades para el cálculo de este coeficiente (Blondel et al., 2004).

Una medida propuesta en (Callaway et al., 2001) define a la asortatividad como una función de correlación en términos de los grados de la red. Esta función vale cero cuando no existe asortatividad y los valores positivos indican presencia de asortatividad, mientras que la negativa indica disasortatividad.

Para redes no dirigidas la asortatividad se calcula mediante el coeficiente de correlación de Pearson del coeficiente de grado k entre dos pares de nodos

[h] Pares	σ_{st}	$\sigma_{st}(n)$	$\sigma_{st}(n)/\sigma_{st}$
(A,B)	1	0	0
(A,D)	1	1	1
(A,E)	1	1	1
(A,F)	1	1	1
(B,D)	1	1	1
(B,E)	1	1	1
(B,F)	1	1	1
(D,E)	1	0	0
(D,F)	1	0	0
(E,F)	1	0	0

Tabla 2.2: Cálculo de coeficiente de intermediación para el nodo C 2.11 de la red mostrada en la figura 2.10.

conectados y se define de la siguiente manera:

$$r = \frac{1}{\sigma_q^2} \left[\sum_{jk} jk(e_{j,k} - q_j q_k) \right] \quad (2.17)$$

donde r es el coeficiente de asortatividad de la red, i, j dos nodos cualesquiera, $e_{j,k}$ es la distribución de probabilidad conjunta de los grados k de dos pares de nodos elegidos al azar. σ_q es la desviación estándar de la distribución de grado de la red y q_k es la distribución de grado que captura el número de enlaces que salen del nodo que no estén conectando a cada par de nodos.

Parte II

Metodologías

En esta sección se definirán los conceptos básicos de redes sociales *online* y presenciales utilizados a lo largo de la Parte 3 de esta tesis. Además se definen los conceptos específicos de Actividad, Respuesta que da el Sistema, Centralidad y Eficiencia que se utilizan a lo largo de esta tesis.

Capítulo 3

Redes Sociales

En este capítulo se mostrará como se pueden analizar diferentes sistemas sociales, en donde cada uno de sus individuos serán los nodos y los enlaces serán representados por los distintos tipos de relaciones que hay entre ellos.

Un sistema social es un ejemplo paradigmático de sistema complejo. Para estudiarlo, no sólo se debe analizar el comportamiento individual y sus características principales, como el mérito que tiene como individuo que pertenece a una sociedad y los valores que pueda aportar a la misma, sino que también es importante considerar las interacciones que tenga con el resto de los individuos de la red y su capacidad real de relacionarse con el resto. En relación a esto se han utilizado los conceptos de centralidad de redes, actividad y respuesta de la red como indicadores para poder caracterizar estos grafos

En relación a esto se ha denominado el concepto de topocracia, el cual está relacionado directamente con las medidas de centralidad que tenga el individuo en la red, como los descritos en la sección 2.3.7. Por otro lado la capacidad de un individuo de producir valor a la sociedad se denominará meritocracia (Borondo et al., 2014) y (Borondo, 2015).

En la representación como grafo o red de un sistema social (que desde ahora se denominará red social), por lo general diremos que los nodos son los individuos que conforman la sociedad, sin embargo es posible definir tantas redes como posibles relaciones existan entre ellos, como por ejemplo redes de amistad, trabajo, comerciales, entre otras. El interés de esta tesis consiste en dos tipos de relaciones muy diferentes, las que se producen en la sociedad digital gracias a la expansión de internet y aquellas que tienen que ver con relaciones de proximidad sin que sea necesario la utilización de aplicaciones *online* que marcan las reglas de las relaciones.

Las redes sociales *online* permiten establecer relaciones entre individuos que no tienen porque conocerse previamente, incluso relaciones individuo-sistema, ya que se pueden enviar mensajes sin un destinatario concreto, para que lleguen a todos los actores del sistema. Evidentemente estas redes son

una parte fundamental de la globalización que se vive actualmente en este primer cuarto del siglo XXI y dentro de los grandes ejemplos de redes sociales *online* tenemos a Twitter (que será utilizado en profundidad en este trabajo) y Facebook.

Por otro lado, las redes sociales de proximidad o presenciales implican una relación bidireccional individuo-individuo más clara. Esto no quiere decir que en estas redes no se utilizan las tecnologías de la comunicación, sino que las que se utilizan son fundamentalmente aquellas que permiten relaciones personales. Las llamadas redes telefónicas o de *email* son un ejemplo de esto, sin olvidar las redes de proximidad «clásicas», como pueden ser las de amistad, como es el caso de la red de Karate estudiada por Zachary (Zachary, 1977) o la de contactos sexuales (Ergün, 2001).

En este trabajo se analizarán las relaciones que se producen entre científicos cuando realizan publicaciones conjuntas o dicho de otra forma, la red de colaboración científica entre los autores, lo que implica relaciones personales de proximidad para obtener resultados publicables.

Como se ha mencionado, Twitter y las colaboraciones científicas son los ejemplos de redes sociales *online* y de proximidad que se estudiarán en esta tesis. A continuación se describirán brevemente sus principales características.

3.1. Twitter

Twitter es un servicio de *microblogging* que permite a los usuarios enviar mensajes de texto plano de corta longitud, de un máximo de 280 caracteres, a los que se les denomina *tweets*. Dentro de estos mensajes se pueden añadir diversas «etiquetas», como por ejemplo los *hashtags*, que son palabras claves que permiten que el sistema logre identificar de forma automática el tema que está tratando el mensaje. Para escribir un *hashtag* se debe anteponer una almohadilla (#) antes de cada palabra (Twitter, 2011). Análogamente se pueden mencionar y contestar a otros usuarios escribiendo el nombre de éste antes de un arroba (@) y finalmente es posible retuitear un mensaje, que es la opción de compartir o volver a enviar el mensaje de otro usuario, colocando las letras «RT» antes del mensaje.

Una de las grandes ventajas de esta red social es que es posible hacer uso de su API (*Application Programming Interface*), la cual es un servicio que provee funciones para hacer uso de un servicio web ajeno para uso personal (Hipertextual, 2014).

A través de la API de Twitter es posible obtener información de cuentas de usuarios, *tweets*, respuestas de mensajes, junto a una gran colección de metadatos que contiene información relevante del *tweet*. En la tabla 3.1 se observa la diferencia entre los distintos tipos de servicios de API que ofrece Twitter.

Tweets	Estándar	Premium	Empresarial
Publicar	Sí		
Buscar Tweets: 7-días	Si		
Buscar Tweets: 30-días		Si	Si
Buscar Tweets: Todos		Si	Si
Filtrar Tweets	Si		Si
Muestra de Tweets	Si		Si
Descargar Tweets masivamente			Si

Tabla 3.1: Diferencia de las distintas versiones de la API de Twitter. La versión estándar es gratuita.

Dentro de los ejemplos de metadatos que se encuentran en Twitter se incluyen los filtros de fechas, relación entre usuarios, geolocalización, número de menciones del *tweet*, *hashtags* utilizados, entre otros.

3.2. Colaboración Científica

Durante esta tesis la colaboración científica se construye por medio de la descarga de publicaciones científicas a través de la API de Thomson Reuters, incluyendo todos sus metadatos (figura 3.1), que incluyen los nombres de los autores, nombre de las instituciones asociadas, disciplinas científicas, países involucrados, año de publicación, citas de la publicación, entre otros.

La forma de construcción de este tipo de redes se tratará en profundidad en el capítulo 5, sin embargo es importante mencionar que, debido a la estructura de los datos, es posible construir redes derivadas que pueden no considerarse como sociales, ya que los nodos no son representados como individuos, pero a la vez pueden ayudar a entender diferentes sistemas sociales, como económicos, productivos, etc.

Debido a la naturaleza de las redes sociales *online*, en las que muchas de las acciones de los individuos quedan registrados, produciéndose una traza digital de la actividad individual, mediante técnicas de *big data* se puede construir redes de gran tamaño, con mucha información que permite realizar análisis profundos del comportamiento humano (Olivares et al., 2019), (Morales et al., 2015a), (Borondo et al., 2012), (Herrera-Yagüe et al., 2015), (Herdagdelen et al., 2013) y (Martin-Gutierrez et al., 2018), sin embargo, esto no es tan fácil cuando se pretende analizar las redes de proximidad, ya que la obtención de datos requiere un trabajo mas complejo.

En los últimos años se han realizado estudios a partir de datos de compañías telefónicas (Martin-Gutierrez et al., 2016), (Morales et al., 2015b) y (Creixell et al., 2012). Conseguir estos datos no siempre es fácil y pocas veces se posee un número de datos comparable a los de Twitter y tampoco es

sencillo observar sus evoluciones temporales.

Figura 3.1: Estructura de una publicación científica indexada en la Web of Science de Thomson Reuters.

En esta tesis se ha conseguido construir redes sociales de proximidad de tamaño similar a las que se utilizan en los trabajos de redes sociales *online*, con el valor añadido de tener datos a lo largo de diferentes años, lo que permite realizar un análisis de la evolución temporal del sistema.

3.3. Actividad, Respuesta del Sistema, Eficiencia y Centralidad

Se definirá como **Actividad** de un nodo a una medida individual y cuantificable que depende directamente del individuo y que está relacionado al número de acciones que es capaz de generar en un período de tiempo dado. Un ejemplo de esto son el número de publicaciones científicas de un autor, como también puede ser el número de tuits emitidos por un usuario, dentro de un período determinado de tiempo.

El concepto de **Respuesta del Sistema** se define como una medida cuantificable del efecto que se produce en la red debido a la actividad de un nodo. Para el caso de las redes de colaboración científica esto puede

determinarse a través de las citas que recibe la publicación de un autor, mientras que para las redes sociales *online* esta medida puede obtenerse a través del número de *retweets* de un usuario.

Existen otras características que dependen de la respuesta que da el sistema a la actividad que realiza un nodo. Para cuantificarlo se define la **Eficiencia** de un nodo como las respuestas que da el sistema a sus acciones individuales.

Por otro lado, la **Centralidad** son las características topológicas de un nodo de una red que le dan ventaja competitiva respecto a otros nodos. No depende de él, sino de las relaciones que tiene con sus vecinos. Por ejemplo, una ciudad comunicada con una buena red de transporte es topológicamente más competitiva que una con un solo acceso. En los próximos capítulos se han definido diferentes medidas topológicas de los nodos de una red, de acuerdo al caso de estudio. Dependiendo del área del conocimiento en la que se esté trabajando, se puede considerar la Centralidad como un tipo de Topocracia.

Por tanto, la Actividad y la Respuesta que da el Sistema son cualidades individuales del nodo y no interviene en la definición de la red, mientras que la centralidad depende solo de la posición que ocupa el nodo en la red. No está clara la correlación que pudiera haber entre ambas características. En ciertos sistemas se podría pensar que un nodo con altas cualidades tiene más posibilidades de estar más conectado (una ciudad productiva tiene más recursos económicos para mejorar sus infraestructuras de comunicación). Sin embargo, es sabido, por ejemplo, que es más fácil encontrar trabajo a través de los contactos sociales que solo enviado los currículos. En el capítulo 4 se analizará esta correlación en el caso de las colaboraciones científicas.

3.4. Características de las redes estudiadas

Uno de los objetivos fundamentales de esta tesis tiene que ver con comprender el fenómeno de las relaciones entre individuos de una población como una característica necesaria para definir a aquellos individuos como parte de un sistema social. Dicho en palabras poco rigurosas, esas relaciones sólo ocurren en el dominio social de aquellos individuos. Así, las relaciones sociales *online*, o virtuales, son una proyección de aquella naturaleza social de los humanos en una dimensión digital, creada por ellos mismos (como sistema) tal vez con la finalidad de derribar barreras que restringen las relaciones tradicionales aumentando así la complejidad del sistema del cual forman parte.

Dada la superación de aquellas restricciones que limitan las relaciones humanas tradicionales, la naturaleza de los enlaces en el mundo digital es completamente distinta a la observada hasta antes de la aparición de la tecnología que permitió mantener, potenciar e incluso crear nuevos tipos de relación antes impensadas. Por esta razón, no sería extraño encontrar nuevas propiedades que emergen en estas redes sociales virtuales como producto

Tipo de red	Coefficiente de Agrupamiento	Coefficiente de Asortatividad
Retweet Chile	0.003790	-0.1738627
Retweet Brasil	0.023735	-0.1141201
Co-autoría Chile	0.935660	0.9948616
Co-autoría Brasil	0.897278	0.8381598

Tabla 3.2: Diferencias entre redes sociales virtuales (elecciones presidenciales de Chile del 2017 y las elecciones presidenciales de Brasil del 2018) y redes sociales presenciales para (co-autoría de Chile 2017 y co-autoría de Brasil 2018).

de nuevos tipos de relaciones humanas donde la identidad, proximidad, popularidad, entre otros conceptos, adquieren nuevas valoraciones.

Por ejemplo, si se compara una red social *online* como aquella que ocurre entre usuarios de Twitter y una con una más tradicional, con una componente presencial, como es la red de colaboración científica, se pueden observar algunas diferencias en sus estructuras de red, tal como se muestra en la tabla 3.2, en donde se aprecia que para las redes virtuales existe una disasortatividad o correlación negativa de los nodos, es decir, la preferencia de conexión no es en base a similitud y tampoco hay tendencia de formar cliques. Por otro lado las redes sociales presenciales poseen una alta tendencia de conexión a nodos similares y además tienen tendencia a generar cliques. En los próximos capítulos se estudiará en profundidad las características adicionales de cada uno de este tipo de redes.

Parte III

Resultados

Esta sección se presentan los resultados de las metodologías propuestas para el análisis de la colaboración científica a través de redes complejas y a su derivado de redes funcionales denominados redes de complemento disciplinar. Adicionalmente, se expande este estudio a las redes sociales *online*, específicamente al estudio de la polarización política en Twitter a través de redes de *Retweets* para las elecciones presidenciales chilenas del 2017.

Capítulo 4

Redes sociales presenciales: Redes de coautoría

En este capítulo se realizará un análisis para redes sociales presenciales, utilizando la base de datos de la Web of Science, descrita en el capítulo 4, cuya estructura se puede ver en la figura 4.1, en donde se marcan los datos necesarios para la construcción de este tipo de redes.

El objetivo de este capítulo es verificar si existe una correlación entre las medidas de Centralidad, Actividad y Respuesta del Sistema de los actores dentro de una red social, que puedan ser utilizados para el desarrollo de indicadores y algoritmos que caractericen diversos sistemas sociales.

Figura 4.1: Estructura de metadatos de una publicación indexada en WoS para la construcción de redes de coautoría.

Research Article

Opinion Polarization during a Dichotomous Electoral Process

G. Olivares,^{1,2} J. P. Cárdenas,² J. C. Losada,¹ and J. Borondo^{1,2}

¹Grupo de Sistemas Complejos, Escuela Técnica Superior de Ingeniería Agronómica, Alimentaria y de Biosistemas, Universidad Politécnica de Madrid, Avda. Complutense s/n, 28040 Madrid, Spain

²Net-Works, Angamos 451, Rehaca, Viña del Mar, Chile

Figura 4.2: Encabezado de un artículo científico, firmado por G. Olivares, J.P. Cárdenas, J.C. Losada y J. Borondo.

4.1. Introducción

Se definirá como red de coautoría a una red social presencial y no dirigida y pesada, en donde dos autores estarán conectados entre sí en el caso que ambos hayan firmado juntos una misma publicación científica. Para ejemplificar mejor la confección de esta red, tomemos como referencia la cabecera del artículo que se observa en la figura 4.2.

En la tercera fila del artículo de la figura 4.2 se presentan los nombres de los co-autores de la publicación, por lo que para realizar la red de co-autoría de este artículo científico se generan enlaces entre todos ellos.

De la misma forma a lo anterior, se pueden seguir sumando publicaciones y vínculos entre sus co autores, obteniendo estructuras de red aún más grandes y complejas, tal como se aprecia en la figura 4.3.

Figura 4.3: Ejemplo de una red de co-autoría para tres autores chilenos en el período 2008-2018, cuyos nodos son presentados en color rojo. Los nodos representan los autores, su tamaño el número de publicaciones realizadas y el color la institución principal a la cual está afiliado. Los enlaces indican la co-autoría entre dos autores y el tamaño del enlace el número de publicaciones en las que han sido co-autores.

4.2. Metodología

Uno de los grandes problemas a la hora de generar relaciones de coautoría es la cantidad de firmas distintas que tiene un mismo autor. Dentro de las soluciones propuestas existen alternativas como ORCID (ORCID, 2018) que es un código alfanumérico que identifica de manera única a científicos y otros autores académicos. El problema de este código es que necesariamente debe ser entregado al momento en que la publicación fue aceptada por la revista científica, dejando muchas de las publicaciones más antiguas sin normalizar.

En Chile se ha ido desarrollando otra solución. CONICYT, una institución chilena que depende directamente del Ministerio de Educación y que además se encarga de desarrollar, promover y fomentar la ciencia y la tecnología en Chile, desde hace varios años ha estado realizando una desambiguación de autores de forma manual. Esto significa que un equipo de trabajadores se ha dedicado a asignar cada una de las distintas firmas a un único autor, cuando corresponda, pero sólo a autores chilenos.

El tema de la desambiguación de autores es muy importante, ya que de esta forma la precisión de los resultados aumenta, especialmente cuando se quiere calcular las cualidades del nodo, en las que encontramos Centralidad, Actividad, Respuesta del Sistema y Eficiencia. Por esta razón se decidió utilizar la base de datos de la Web of Science, en donde se han descargado un total de 93626 publicaciones que tengan al menos un autor con afiliación chilena en el período comprendido entre el año 2008 y 2018. Además, de estas publicaciones se obtiene un total de 67553 autores únicos chilenos, ya normalizados.

Por otro lado, existe otro problema al momento de generar las redes, que es el exceso de autores que pueden tener algunas publicaciones, pudiendo superar fácilmente los 1000 autores en casos de publicaciones indexadas en revistas de la disciplina *Astronomy & Astrophysics*. Este tipo de casos merman por completo el rendimiento de cálculo de medidas topológicas de la red y para solucionarlo se decidió considerar sólo aquellas publicaciones que no superen 100 autores.

Finalmente se procedió a utilizar la clasificación OECD de áreas científicas, la que consiste en un esquema de distribución del conocimiento para clasificar las unidades de ejecución de investigación y desarrollo experimental, en las que incluimos a *Agricultural Sciences*, *Medical and Health Sciences*, *Engineering and Technology*, *Social Sciences*, *Humanities* y *Natural Sciences*. Éstas áreas pueden homologarse con las disciplinas científicas de WoS y su tabla de conversión se presenta en el apéndice C.

Se procedió a generar una red de co-autoría para cada una de estas seis áreas de la OECD y se eligió total de autores chilenos previamente desambiguados (y que tuvieran vínculos de co-autoría), generando además las relaciones de sus vecinos directos dentro de esa área. Adicionalmente, del total

de autores para cada red, sólo se consideraron aquellos que se encuentran dentro de la componente gigante, de manera tal que sus medidas de centralidad puedan ser comparables entre sí. En la tabla 4.1 se presenta el número de autores totales de la red compleja y el porcentaje de ellos que se encuentran dentro de la componente gigante.

Red de coautoría para cada área de la OECD	Total de autores	Porcentaje en la CG
Agricultural Sciences	1404	80 %
Medical and Health Sciences	4278	85 %
Engineering and Technology	2361	71 %
Social Sciences	1335	26 %
Humanities	598	8 %
Natural Sciences	5348	85 %

Tabla 4.1: Total de autores chilenos con al menos una colaboración científica para cada una de las redes de co-autoría de cada área de la OECD. A la derecha se presenta el porcentaje de autores chilenos que pertenece a la componente gigante.

4.3. Actividad, Respuesta del Sistema y Centralidad en las redes de coautoría

Una vez construidas las redes de co-autoría, se procede a estudiar sus propiedades de Centralidad para correlacionarlas con su correspondiente actividad y Respuesta del Sistema. Como se mencionó anteriormente, la red de co-autoría es una red pesada no dirigida, en la que el grado de un nodo refleja el número de investigadores con los que ha colaborado. Hay que tener en cuenta que un artículo puede estar firmado por muchos autores, por lo que un investigador incluso con un artículo con 100 firmantes, tendrá un grado $k = 100$, pero una producción muy baja. Las medidas de Centralidad analizadas son el Pagerank y el grado, descritas en el capítulo 2. La Actividad será caracterizada por su actividad (número de publicaciones) y la Respuesta del Sistema estará dada por su número de citas y su Eficiencia, definida como el número medio de citas por artículo, la cual se observa en la siguiente ecuación.

$$ef_i = \frac{c_i}{p_i} \quad (4.1)$$

donde p_i es el número de publicaciones del autor i y c el número de citas del autor i . En la figura 4.4 se puede observar la distribución de la eficiencia de todos los autores, en donde se muestra que la gran mayoría de los autores posee una muy baja eficiencia y muy pocos autores poseen una eficiencia alta, ajustándose con una ley de potencia.

Se han analizado estas tres características de cada investigador, porque la actividad no necesariamente es una medida de la calidad de la investigación, ya que puede depender de otros factores, como la pertenencia a un grupo grande, con medios técnicos importantes o a su capacidad de conseguir financiación, sin embargo, el número de citas de un artículo es una medida objetiva de la calidad del mismo.

En la figura 4.5 se muestra la correlación entre la actividad de cada investigador con su Pagerank. Cada punto corresponde a un investigador. Para cuantificar la correlación se ha utilizado el coeficiente de correlación de Pearson r (Kendall y Stuart, 1973). En esta figura se aprecia que para todas las áreas de la OECD el coeficiente r entre el Pagerank y el número de publicaciones es de al menos 0.92 indicando una buena correlación, sin embargo para *Natural Sciences* decae fuertemente.

Siguiendo la misma metodología anterior, la figura 4.6 muestra la correlación r entre el coeficiente de grado y las publicaciones, en donde el coeficiente r es de al menos 0.606 si no se considera *Natural Sciences*.

Figura 4.4: Distribución de eficiencia de los autores en escala logarítmica.

Si se analiza en profundidad el caso de *Natural Sciences*, para ambas correlaciones anteriores se logran observar una serie de 11 puntos extremos correspondientes a once autores con mas de 200 publicaciones. Si se eliminan a estos autores se obtienen los resultados de la figura 4.7, cuyo coeficiente r es de 0.841 para el número de publicaciones vs Pagerank y de 0.736 para

el número de publicaciones y el coeficiente de grado, lo que indicaría una correlación entre ambas medidas, exceptuando los valores extremos. Debido a que el conjunto de datos extremos es muy pequeño, no se puede realizar un análisis profundo de estos casos, sin embargo, estos autores en su mayoría son pertenecientes a la disciplinas relacionadas a Ciencias Físicas y Astronomía, cuyas publicaciones en muchos casos pueden superar las mil co-autorías.

Figura 4.5: Diagrama de Dispersión y correlación entre Pagerank y número de publicaciones de los autores en la red, correspondientes a cada una de las seis áreas de la OECD.

Figura 4.6: Diagrama de Dispersión y correlación entre Grado y número de publicaciones de los autores en la red, correspondientes a cada una de las seis áreas de la OECD.

Figura 4.7: Diagrama de Dispersión y correlación entre Pagerank y número de publicaciones (izquierda) y entre Grado y número de publicaciones (derecha) para el área *Natural Sciences* eliminando los 11 autores con más de 200 publicaciones.

Una vez visto que la medida de centralidad más importante es el Pagerank, en la figura 4.8 se muestra su correlación con el número de citas de cada investigador por área del conocimiento, cuyos resultados muestran que a pesar de que se observa una correlación para *Agricultural Sciences*, *Medical and Health Sciences*, *Social Sciences* y *Humanities*, no sucede lo mismo para *Natural Sciences* y *Engineering and Technology*. Sin embargo la correlación se da mas que nada por los puntos extremos, los cuales si son eliminados eliminan toda correlación entre el Pagerank y las citas.

Finalmente se realizó la correlación entre el Pagerank y la Eficiencia (ver figura 4.9), en donde no se observa ninguna correlación entre los valores.

Figura 4.8: Diagrama de Dispersión y correlación entre Pagerank y número de citas de los autores en la red, correspondientes a cada una de las seis áreas de la OECD.

Figura 4.9: Diagrama de Dispersión y correlación entre Pagerank y Eficiencia de los autores en la red, correspondientes a cada una de las seis áreas de la OECD.

4.4. Discusión

De los resultados presentados se observa que tanto el Pagerank y el Degree se correlacionan con el número de publicaciones del autor, sin embargo no se observa ninguna correlación entre las medidas topocráticas y el número de citas o incluso la eficiencia.

El Pagerank es una medida de centralidad que depende de la importancia de los vecinos con la cual está conectado, es decir, depende de la posición global que se ocupa, mientras que el grado es una medida que sólo depende de los primeros vecinos, pudiendo existir nodos con muchos enlaces, pero muy aislados de la red global.

Entre estas dos medidas se ha encontrado que el Pagerank tiene mayor correlación con la actividad que el grado, sin embargo, debido a que el Pagerank es un método recursivo, el grado puede resultar mucho más útil en

redes que se requieren cálculos en tiempo real, por la inmediatez del resultado. Por otro lado, cuando mayor actividad tenga un investigador, mayor probabilidad tienen de ser un nodo central en la red de coautorías a la que pertenece, que incluye a segundos y terceros vecinos (como las mostradas en la figura 4.3), sin embargo, esto no es tan evidente cuando se relaciona a la actividad con el número de coautores directos.

Un caso particular de estudio es el caso de *Natural Sciences*, ya que el coeficiente de correlación de Pearson no superaba un valor de 0.2. Sin embargo, al eliminar once autores con más de 200 publicaciones nuevamente se obtiene una alta correlación tanto con el Pagerank y el coeficiente de grado. Esto podría significar que existe un cambio de estado cuando se encuentran autores con una producción de artículos científicos muy alta, cuyo caso de estudio no puede ser representado de manera óptima para Chile debido a la poca cantidad de autores con este número de publicaciones.

Finalmente, se observa que no existe una correlación entre la meritocracia (medida a través del número de citas o por la eficiencia del autor) y la topocracia (medida a través del Pagerank o el coeficiente de grado) de los autores, lo que podría significar que una alta eficiencia no depende de la posición estratégica de la red ni a la cantidad de publicaciones de un autor, sino que más bien el número de citas tiene relación a otros factores que no son capturados a la hora de construir las redes.

Capítulo 5

Redes de complemento disciplinar

Este capítulo propone establecer una metodología de representación de estructuras disciplinares, a las que llamamos redes de complemento disciplinar, las cuales son una forma distinta de representar las redes físicas, debido a que los metadatos que poseen las publicaciones permiten la transformación de los nodos (autores) a un concepto abstracto que son las disciplinas.

Este capítulo parte de la hipótesis en que se puede generar un factor universal para medir la importancia de una disciplina a través de su Centralidad, Actividad y Respuesta del Sistema, y que existe una diferencia estructural en las redes de complemento disciplinar entre los países desarrollados y los países en vías de desarrollo.

El objetivo principal es la de poder comparar entre países, cómo se comporta la ciencia en cada uno de ellos. Por otro lado también se plantea una ecuación para describir el desarrollo de la ciencia en cada uno de estos países.

5.1. Introducción

Se definirá como red de complemento disciplinar a una red de disciplinas pesada $G(D, E)$, donde D son el conjunto de disciplinas distintas y E sus relaciones de co-ocurrencia en artículos científicos. Un ejemplo de construcción de la red se puede observar en la figura 5.2.

Para la confección de este tipo de redes se utilizarán exclusivamente las publicaciones científicas indexadas en la base de datos de Web of Science, descritas en el capítulo 4. Como ya se ha mencionado, esta base de datos permite identificar a los investigadores los artículos que han sido citados y además, ver de una forma muy detallada cada uno de los metadatos de cada publicación.

En esta ocasión se utilizará la parte de la derecha de la figura 3.1 descrita

en el capítulo 4, la cual se detalla en la siguiente figura 5.1.

Figura 5.1: Diagrama de datos de una publicación científica. Se debe considerar que esta parte sólo incluye la parte de las revistas y áreas científicas.

En este caso se trabajará con las disciplinas científicas asociadas a cada publicación. Se dirá que un artículo científico está asociado a una o más disciplinas científicas, dependiendo de la(s) disciplina(s) científica(s) a la cual esté asociada la revista científica en donde fue publicado el artículo. Por esta razón es posible relacionar a dos disciplinas cuando participan en conjunto en una publicación a través de una revista.

Figura 5.2: Ejemplo de generación de una red de complemento disciplinar para tres publicaciones distintas.

Posteriormente, al igual que las redes de autores, se obtiene una estructura completa de red de complemento disciplinar cuando añadimos más publicaciones, cuyo ejemplo se puede observar en la figura 5.3.

Figura 5.3: Ejemplo de red de complemento disciplinar generada para las publicaciones de tres autores chilenos. Cada nodo representa una disciplina científica, el color representa un Área de la OECD y el tamaño el número de publicaciones desarrollada en cada disciplina. Los enlaces representan los vínculos entre dos disciplinas científicas y el tamaño del enlace el número de veces que ambas disciplinas interaccionaron.

5.2. Metodología para diferenciar estructuras disciplinares entre países

Utilizando la metodología de redes de complemento disciplinar, se procedió a analizar los vínculos interdisciplinarios de los siguientes países: Argentina, Brasil, Chile, Colombia, Cuba, Finlandia, Grecia, Islandia, Israel, Nueva Zelanda, Polonia, Portugal, España y Venezuela, a través de los datos obtenidos a través de la base de datos previamente mencionada, en el período comprendido entre el 2008-2018. En la tabla 5.1 se observan el número de publicaciones descargados para cada uno de los países señalados.

La elección de estos países radica principalmente en la diversidad eco-

País	Publicaciones
Argentina	89070
Brasil	406494
Chile	68375
Colombia	41106
Cuba	10570
Finlandia	118225
Grecia	106242
Islandia	9623
Israel	133219
Nueva Zelanda	86876
Polonia	233674
Portugal	122850
España	524773
Venezuela	12488

Tabla 5.1: Número de publicaciones para cada uno de los países estudiados en el período 2008-2018.

nómica y nivel de desarrollo que representa cada uno. Según el Fondo Económico Internacional (Internacional, 2019), dentro del conjunto de países seleccionados, seis de ellos pertenecen a economías desarrolladas (Finlandia, Grecia, Islandia, Israel, Nueva Zelanda, Portugal y España), otros dos serán considerados desarrollados a futuro (Chile y Polonia) y el resto de los cinco países restantes se clasifican como países emergentes (Argentina, Brasil, Colombia, Cuba y Venezuela).

Por otro lado, en el apéndice C se muestra una tabla de conversión de disciplinas, las cuales se han dividido en tres niveles.

El primer nivel corresponde a un conjunto de 251 disciplinas clasificadas por WoS. Este nivel puede ser agrupado en otro conjunto de 38 sub categorías y finalmente se tiene el último subconjunto correspondiente a 6 grandes áreas de la OECD, en donde se encuentran *Natural Sciences*, *Medical and Health Sciences*, *Engineering and Technology*, *Agricultural Sciences*, *Social Sciences* y *Humanities*.

La base de la visualización es un diseño circular (Becker y Rojas, 2001), que consiste en una estructura de visualización en donde los nodos son colocados dentro de un círculo. Para este caso, la visualización tiene la particularidad que se generan 6 diseños circulares dentro de la misma visualización, en donde se agrupan todas las disciplinas dentro de cada círculo correspondiente a su área de la OECD.

Respecto al tamaño del nodo, este será proporcional a la eficiencia de la disciplina, cuya métrica fue definida en el capítulo 4 para los autores, mientras que el tamaño del enlace corresponde al número de relaciones entre dos disciplinas. Para comparar entre distintos países, estos valores fueron normalizados por el máximo valor, a modo de estandarizar en una misma escala a todos los países. En el apéndice B se presentan las visualizaciones

desarrolladas para cada uno de los 14 países estudiados, indicando además, el nombre para cada uno de los nodos dentro de la red.

5.3. Construcción del conocimiento científico

Uno de los grandes desafíos de la economía es reducir la pobreza a nivel mundial, sin embargo, dada la complejidad que existe, especialmente a las estructuras internas de cada nación, la fórmula que sirve para un país no necesariamente será equivalente para otro. A pesar de esto, se han desarrollado diversos indicadores económicos que, de alguna forma, logran captar la realidad individual de cada nación. Uno de estos indicadores es el Producto Interno Bruto (PIB), el cual es una magnitud que expresa el valor monetario de la producción de bienes y servicios durante un periodo de tiempo determinado. Como el efecto del número de habitantes puede influir drásticamente en estos números, lo que se suele hacer es dividir esta magnitud por el número de habitantes de cada país, de esta forma se puede observar una métrica comparable para todos.

Otro indicador, algo más moderno, es el Índice de Complejidad Económica (ECI), el cual es una medida de intensidad relativa de conocimiento de una economía, tomando en cuenta cuales son los productos que exportan (Hidalgo y Hausmann, 2009). De esta forma, países que exportan bienes y/o servicios de mayor valor agregado, como biotecnología, robótica, etc, tendrán un ECI mas alto, mientras que otros países que se dedican principalmente a la exportación de materias primas, como petróleo, minerales, etc., poseen un ECI más bajo.

Ambos indicadores mencionados previamente han sido validados como medidas económicas relevantes para ayudar a predecir el crecimiento económico futuro (ECI) o para establecer una diferenciación entre países en vías de desarrollo y países desarrollados (PIB).

Por otro lado, uno de los desafíos actuales es encontrar un indicador que permita relacionar fuertemente la academia con la industria, de modo tal que el conocimiento interno de cada nación permita ser un complemento perfecto para el desarrollo del país.

En esta sección se planteará una métrica que, a través de redes de complemento disciplinar, se correlacione fuertemente con el PIB.

5.4. Metodología del índice de complejidad disciplinar

Para este estudio se han considerado los 14 países utilizados para la construcción de las visualizaciones de redes de complemento disciplinar, uti-

lizando nuevamente la base de datos de la Web of Science, en donde se han descargado las publicaciones científicas del período 2008-2018 para cada uno de los países mencionados.

Adicional a lo anterior, se ha realizado una red disciplinar única, incluyendo la totalidad de publicaciones para todos los países mencionados en el mismo período, para describir la importancia de cada disciplina de acuerdo a la siguiente ecuación:

$$FID(d) = C_b(d) * k_d * pub_d * cit_d \quad (5.1)$$

donde $FID(d)$ es el Factor de Importancia Disciplinar para la disciplina d , $C_b(d)$ es el coeficiente de intermediación para la disciplina d y k_d es el coeficiente de grado para la disciplina d . Además, pub_d es el número de publicaciones de d y cit_d es el número total de citas de la disciplina d .

La ecuación 5.1 plantea ponderar una medida que incorpore la multidisciplinariedad, la Actividad y la Respuesta del Sistema hacia la disciplina, en donde se recompensa a aquellas disciplinas que posean un mayor número de enlaces k y un coeficiente de intermediación mayor, y además se recompensa a aquellas disciplinas que poseen un número mayor de publicaciones y citas.

En la tabla 5.2 se presenta el FID para cada una de las disciplinas presentadas dentro de la red.

Disciplina	FID	Disciplina	FID
Biochemistry & Molecular Biology	1.693	Hematology	6.00E-05
Environmental Sciences	1.3515	Psychology, Applied	6.00E-05
Chemistry, Physical	0.9058	Language & Linguistics	6.00E-05
Materials Science, Multidisciplinary	0.8168	Reproductive Biology	5.00E-05
Engineering, Electrical & Electronic	0.3942	Philosophy	5.00E-05
Neurosciences	0.2808	Psychology, Social	5.00E-05
Public, Environmental & Occupational Health	0.2551	Parasitology	5.00E-05
Pharmacology & Pharmacy	0.2007	Mycology	5.00E-05
Biotechnology & Applied Microbiology	0.1713	Archaeology	5.00E-05
Chemistry, Multidisciplinary	0.1546	Imaging Science & Photographic Technology	5.00E-05
Ecology	0.133	Transportation	4.00E-05
Computer Science, Interdisciplinary Applications	0.1041	Engineering, Manufacturing	4.00E-05

Disciplina	FID	Disciplina	FID
Plant Sciences	0.0975	Gastroenterology & Hepatology	4.00E-05
Engineering, Chemical	0.0792	Respiratory System	4.00E-05
Energy & Fuels	0.078	Law	4.00E-05
Immunology	0.075	Materials Science, Biomaterials	4.00E-05
Oncology	0.0595	Social Sciences, Biomedical	3.00E-05
Surgery	0.0594	Orthopedics	3.00E-05
Geosciences, Multidisciplinary	0.0555	Transportation Science & Technology	3.00E-05
Economics	0.0518	Information Science & Library Science	3.00E-05
Genetics & Heredity	0.0452	Spectroscopy	3.00E-05
Food Science & Technology	0.0404	Ophthalmology	3.00E-05
Cell Biology	0.0385	Paleontology	3.00E-05
Psychiatry	0.0359	Planning & Development	2.00E-05
Multidisciplinary Sciences	0.0334	Psychology, Clinical	2.00E-05
Chemistry, Analytical	0.0263	Agricultural Engineering	2.00E-05
Zoology	0.0185	Evolutionary Biology	2.00E-05
Medicine, General & Internal	0.0175	Education, Scientific Disciplines	2.00E-05
Engineering, Civil	0.0152	Geography	2.00E-05
Education & Educational Research	0.0149	Transplantation	2.00E-05
Clinical Neurology	0.0115	Crystallography	2.00E-05
Radiology, Nuclear Medicine & Medical Imaging	0.0108	Religion	2.00E-05
Mathematics, Applied	0.0088	Remote Sensing	2.00E-05
Pediatrics	0.0087	Engineering, Industrial	2.00E-05
Physics, Applied	0.0087	Otorhinolaryngology	2.00E-05
Astronomy & Astrophysics	0.0076	Metallurgy & Metallurgical Engineering	2.00E-05
Microbiology	0.0076	Critical Care Medicine	1.00E-05
Physics, Multidisciplinary	0.0069	Medical Informatics	1.00E-05
Nutrition & Dietetics	0.0066	Rheumatology	1.00E-05
Chemistry, Applied	0.0064	Geology	1.00E-05

Disciplina	FID	Disciplina	FID
Computer Science, Artificial Intelligence	0.006	Health Policy & Services	1.00E-05
Biology	0.006	Mining & Mineral Processing	9.00E-06
Nanoscience & Nanotechnology	0.0058	Area Studies	9.00E-06
Infectious Diseases	0.0056	Computer Science, Hardware & Architecture	9.00E-06
Environmental Studies	0.0049	Materials Science, Composites	7.00E-06
Polymer Science	0.0049	GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY	7.00E-06
Marine & Freshwater Biology	0.0047	Humanities, Multidisciplinary	6.00E-06
Mechanics	0.0047	Ergonomics	6.00E-06
Medicine, Research & Experimental	0.0045	Family Studies	5.00E-06
Engineering, Biomedical	0.0034	Art	5.00E-06
Engineering, Mechanical	0.0033	Ethics	5.00E-06
Instruments & Instrumentation	0.0032	Soil Science	5.00E-06
Agronomy	0.0029	Cell & Tissue Engineering	5.00E-06
Biochemical Research Methods	0.0027	Engineering, Aerospace	5.00E-06
Toxicology	0.0027	Materials Science, Characterization & Testing	5.00E-06
Optics	0.0027	Medical Laboratory Technology	5.00E-06
Sport Sciences	0.0026	Virology	4.00E-06
Veterinary Sciences	0.0025	Mineralogy	4.00E-06
Operations Research & Management Science	0.0025	Microscopy	4.00E-06
Physics, Particles & Fields	0.0022	Audiology & Speech-Language Pathology	4.00E-06
Obstetrics & Gynecology	0.0021	Ornithology	4.00E-06
Management	0.0019	Developmental Biology	3.00E-06

Disciplina	FID	Disciplina	FID
Nuclear Science & Technology	0.0018	Literature	3.00E-06
Mathematics, Interdisciplinary Applications	0.0015	Architecture	3.00E-06
Physics, Atomic, Molecular & Chemical	0.0015	Criminology & Penology	3.00E-06
Cardiac & Cardiovascular Systems	0.0015	Social Work	2.00E-06
Dentistry, Oral Surgery & Medicine	0.0013	Limnology	2.00E-06
Statistics & Probability	0.0013	Urban Studies	2.00E-06
Computer Science, Information Systems	0.0012	Gerontology	2.00E-06
Meteorology & Atmospheric Sciences	0.0011	Literature, Romance	2.00E-06
Geography, Physical	0.0011	Music	2.00E-06
History	0.001	Women's Studies	2.00E-06
Chemistry, Inorganic & Nuclear	0.001	Social Sciences, Mathematical Methods	2.00E-06
Chemistry, Medicinal	0.001	Psychology, Biological	1.00E-06
Telecommunications	0.0009	Horticulture	1.00E-06
Endocrinology & Metabolism	0.0008	Cultural Studies	1.00E-06
Oceanography	0.0008	Anesthesiology	1.00E-06
Physics, Mathematical	0.0008	History Of Social Sciences	1.00E-06
Agriculture, Multidisciplinary	0.0008	Logic	1.00E-06
Physics, Condensed Matter	0.0007	Medieval & Renaissance Studies	9.00E-07
Physiology	0.0007	Social Issues	8.00E-07
Biophysics	0.0006	Materials Science, Paper & Wood	7.00E-07
Forestry	0.0006	Industrial Relations & Labor	7.00E-07
Engineering, Multidisciplinary	0.0006	Materials Science, Textiles	6.00E-07
Psychology	0.0005	International Relations	6.00E-07
Psychology, Experimental	0.0005	Ethnic Studies	5.00E-07
Computer Science, Theory & Methods	0.0005	Computer Science, Cybernetics	5.00E-07

Disciplina	FID	Disciplina	FID
Fisheries	0.0005	Psychology, Educational	5.00E-07
Social Sciences, Inter-disciplinary	0.0005	Demography	5.00E-07
Water Resources	0.0005	Literary Theory & Criticism	4.00E-07
Rehabilitation	0.0004	Materials Science, Coatings & Films	4.00E-07
Physics, Nuclear	0.0004	Andrology	3.00E-07
Automation & Control Systems	0.0004	Public Administration	3.00E-07
Psychology, Multidisciplinary	0.0003	Medicine, Legal	3.00E-07
Pathology	0.0003	Engineering, Ocean	3.00E-07
Anthropology	0.0003	Allergy	3.00E-07
Geochemistry & Geophysics	0.0002	Psychology, Psychoanalysis	1.00E-07
Business	0.0002	Psychology, Mathematical	1.00E-07
Sociology	0.0002	Asian Studies	1.00E-07
Communication	0.0002	Agricultural Economics & Policy	6.00E-08
Nursing	0.0002	Medical Ethics	2.00E-08
Geriatrics & Gerontology	0.0002	Classics	1.00E-08
Linguistics	0.0002	Theater	1.00E-08
Engineering, Environmental	0.0002	Film, Radio, Television	6.00E-09
Acoustics	0.0002	Literature, British Isles	2.00E-10
Mathematical & Computational Biology	0.0002	Literature, German, Dutch, Scandinavian	4.00E-11
Political Science	0.0002	Literary Reviews	1.00E-11
Peripheral Vascular Disease	0.0002	Folklore	3.00E-12
Health Care Sciences & Services	0.0002	Business, Finance	0.00E+00
Construction & Building Technology	0.0002	Education, Special	0.00E+00
Electrochemistry	0.0001	Emergency Medicine	0.00E+00
Behavioral Sciences	0.0001	Engineering, Geological	0.00E+00
Thermodynamics	0.0001	Engineering, Marine	0.00E+00
Entomology	0.0001	Engineering, Petroleum	0.00E+00

Disciplina	FID	Disciplina	FID
Biodiversity Conservation	0.0001	Integrative & Complementary Medicine	0.00E+00
Psychology, Developmental	0.0001	Literature, African, Australian, Canadian	0.00E+00
Dermatology	0.0001	Literature, American	0.00E+00
Hospitality, Leisure, Sport & Tourism	0.0001	Literature, Slavic	0.00E+00
Urology & Nephrology	0.0001	Materials Science, Ceramics	0.00E+00
Tropical Medicine	0.0001	Neuroimaging	0.00E+00
Mathematics	0.0001	Physics, Fluids & Plasmas	0.00E+00
Agriculture, Dairy & Animal Science	0.0001	Poetry	0.00E+00
Computer Science, Software Engineering	0.0001	Primary Health Care	0.00E+00
History & Philosophy Of Science	0.0001	Robotics	0.00E+00
Chemistry, Organic	0.0001	Substance Abuse	0.00E+00
Anatomy & Morphology	0.0001		

Tabla 5.2: FID para cada una de las disciplinas estudiadas.

En la tabla 5.2 se observa que disciplinas como *Biochemistry & Molecular Biology* es puntuado con nota máxima $FID = 1,693$, lo que quiere decir que, a nivel topocrático y meritocrático es aquella que lidera en la red. Por otro lado, aquellas disciplinas con poco nivel topocrático y meritocrático como *Poetry* son ponderados con un $FID = 0$. Es importante agregar que disciplinas como *Astronomy & Astrophysics*, a pesar de su alto impacto, al no presentar interdisciplinariedad es puntuado con un $FID = 0,0076$.

Con los resultados anteriores, se procedió a generar una medida que permita comparar como se comporta el conocimiento en cada país, la cual se describe en la siguiente ecuación:

$$ICD_p = \sum_d ef_p(d) * FID(d) \quad (5.2)$$

en donde $ICD(p)$ es el Índice de Complejidad Disciplinar para el país p y $ef_p(d)$ es la eficiencia de la disciplina d para el país p .

Con estos resultados se procede a normalizar el indicador a través el método min-max, el que consiste en transformar los valores del ICD a una nueva escala por determinar (Patro y Sahu, 2015) y su ecuación se describe

de la siguiente manera:

$$v' = \frac{v - \min_A}{\max_A - \min_A}(\text{newmax}_A - \text{newmin}_A) + \text{newmin}_A \quad (5.3)$$

donde v' es el nuevo valor transformado, v el antiguo valor, $\min_A$ el mínimo del conjunto de datos A , $\max_A$ el máximo del conjunto de datos A , newmin_A el nuevo mínimo y newmax_A el nuevo máximo. Por lo tanto, finalmente se tiene

$$-1 \leq ICD_p \leq 1 \quad (5.4)$$

5.5. Resultados

Aplicando la ecuación 5.2 para cada uno de los países del estudio, se obtiene un único indicador ICD por país, el cual se observa en la tabla 5.3 y 5.4. Lo primero que se observa es que existe una clara segmentación por el ICD para los países desarrollados y los demás, en donde todos aquellos países con $ICD > 0$ son países desarrollados, a excepción de Portugal, que a pesar de lo anterior, su $ICD = -0,0057$, es decir, muy cercano a cero.

País	ICD
Argentina	-0.5014
Brasil	-0.7646
Chile	-0.4656
Colombia	-0.4818
Cuba	-0.6781
España	0.4057
Finlandia	0.5804
Grecia	0.1618
Israel	0.6928
Islandia	1
Nueva Zelanda	0.4178
Polonia	-1
Portugal	-0.0057
Venezuela	-0.3793

Tabla 5.3: ICD para cada uno de los países del estudio.

Posteriormente se procedió a observar si existe una correlación entre el PIB per cápita, de acuerdo a los valores entregados por el Banco Mundial en el 2018, y el ICD, cuyos resultados indican que efectivamente existe una correlación entre ambas variables con un $r = 0,823$ y un valor $p = 0,0003$ indicando una alta significancia del modelo, tal como se muestra en la figura 5.5.

Adicionalmente se procedió a verificar si la correlación mejora si se compara el ICD con el PIB per cápita efectivo destinado en I+D para cada país, o dicho de otra forma, el porcentaje del PIB per cápita destinado en I+D en dólares, cuyos valores se observan en la tabla 5.4 y el gráfico de dispersión se

Figura 5.4: ICD por país. En rojo se muestran los países en vías de desarrollo, en morado, países que serán desarrollados en el futuro y en azul, países desarrollados de acuerdo a la clasificación del Banco Mundial del 2019.

Figura 5.5: Gráfico de dispersión entre el PIB per cápita para cada país en el año 2018 y su ICD respectivo. Cada punto representa un país.

País	PIB p/c	% PIB p/c I+D	PIB p/c en I+D
Argentina	21528	0.53 %	114.74424
Brasil	16199	1.27 %	205.07934
Chile	25667	0.36 %	92.91454
Colombia	17406	0.24 %	42.47064
Cuba	9900	0.35 %	34.551
España	40289	1.19 %	477.42465
Finlandia	46342	2.75 %	1272.55132
Grecia	29057	1.01 %	292.60399
Israel	37673	2.08 %	1123.01075
Islandia	54121	4.25 %	1601.47923
Nueva Zelanda	40118	1.26 %	506.69034
Polonia	31430	0.97 %	303.2995
Portugal	31964	1.27 %	404.66424
Venezuela	10399	0.12 %	12.37481

Tabla 5.4: PIB per cápita, Porcentaje del PIB per cápita destinado a I+D y PIB per cápita destinado a I+D para cada país.

Figura 5.6: Gráfico de dispersión entre el PIB per cápita efectivo en I+D para cada país del estudio en el año 2018 y su ICD respectivo. Cada punto representa un país.

Finalmente la figura 5.6 indica que la correlación es mayor cuando se compara con el PIB efectivo per cápita destinado a I+D, con un $r = 0.84$ y un valor $p = 0.0003$ indicando alta significancia del modelo cuando se aplica una regresión lineal y un $R^2 = 0.924$ al aplicar una regresión polinómica de grado 2, lo que indica una mejora respecto a la correlación presentada en la figura 5.5

5.6. Discusión

Los conceptos de topocracia y meritocracia descritos en el capítulo 5 y la metodología de visualización de redes de complemento disciplinar han resultado de utilidad para encontrar un indicador que pueda medir la importancia de cada una de las disciplinas a través de un factor llamado *FID* y además poder describir el conocimiento científico de cada país a través de un indicador *ICD*.

Si se observan los valores de *ICD* para países en vías de desarrollo y el de los países desarrollados, se observa una clara tendencia que indica que los países desarrollados, a excepción de Portugal, poseen un $ICD > 0$, mientras que en los países en vías de desarrollo un $ICD < 0$. Lo interesante de los resultados es que un país desarrollado con poca producción científica como Islandia logró el mayor valor de *ICD* de todo el conjunto de datos. Esto da a

entender que el número de publicaciones no es relevante a la hora de generar conocimiento, sino que la importancia radica en realizar publicaciones con alta eficiencia en las disciplinas científicas mas relevantes descrita por el *FID*.

Por otro lado, el *FID* permite obtener un indicador que mide la importancia de desarrollo para cada una de las disciplinas científicas, ordenándolas según su orden de importancia para poder medir el conocimiento científico individual de cada país. La idea general fue favorecer aquellas disciplinas con un mayor índice topocrático y meritocrático, de esta forma, la interdisciplinariedad juega un rol importante a la hora de medir la importancia disciplinar.

Finalmente, al correlacionar el *ICD* de cada país con el PIB per cápita y el PIB per cápita destinado a I+D se observan correlaciones con $r > 0,8$, además se puede ajustar el modelo a una regresión polinomial de grado 2, cuyo R^2 es de 0.924, sin embargo, para mejorar la robustez del modelo, es necesario hacer una comparación con un conjunto mayor de países. En cualquier caso esta medida es un punto de partida para poder enfocar mejor los esfuerzos de cómo generar ciencia, dependiendo del nivel de desarrollo económico de cada país.

Capítulo 6

Modelo predictivo de disciplinas científicas basado en redes de complemento disciplinar.

En la sección anterior se ha descrito una metodología que permite cuantificar el desarrollo científico de un país a través de la información extraída de las redes de complemento disciplinar.

El objetivo de este capítulo es desarrollar un algoritmo basado en redes de complemento disciplinar para pronosticar y guiar a las instituciones a desarrollar nuevas disciplinas científicas. Además se quiere observar el efecto que tiene la Actividad y la Centralidad al momento de la predicción.

6.1. Introducción

La universidad chilena más antigua es la Universidad de Chile, fundada en 1842. A la fecha, existe un total de 61 universidades debidamente reconocidas, sin embargo sólo 8 de ellas fueron fundadas antes de 1981 (Cruz-Coke, 2004). Como la mayoría de las universidades son relativamente recientes, se entiende que la cantidad de conocimiento previo generado es mucho menor al de las universidades más antiguas, por lo que es muy posible que exista un nicho de desarrollo de conocimiento sin explotar en estas instituciones. Por otro lado el potencial de desarrollo de conocimiento de Chile se puede observar en los siguientes casos. En 1996 se volvió el país con mayor productividad por investigador y además presentó las tasas más altas de publicaciones por habitantes respecto a otros países latinoamericanos en 2001 (Koljatic y Silva, 2001). Adicional a esto, Chile supera a varios países en términos de número total de citas y publicaciones por investigador, debido a una gran colabora-

ción internacional (Gonzalez-Brambila et al., 2016). Una de las problemáticas que tiene un país como Chile, es la extrema dependencia de la explotación de materias primas (Lauterbach, 2016), por lo que es sumamente importante la generación de conocimiento aplicado para poder transformar estos elementos a nuevos productos de alto valor y las universidades pueden cumplir un rol fundamental en este sentido, pues como ya se mencionó, Chile ya posee las condiciones necesarias para ello en términos de productividad. Se requiere que las universidades puedan generar competencias en diversas áreas críticas para facilitar la transferencia de conocimiento a los diversos sectores productivos y una posible solución es guiarlas a enfocar sus esfuerzos al desarrollo de estas áreas, enfocándose en todo el conocimiento generado hasta el momento, sin embargo a la fecha no se han encontrado herramientas que faciliten resolver esta problemática. Una de las hipótesis que se plantea es que al plasmar en un grafo el conocimiento histórico de Chile, será posible modelar este conocimiento de manera tal que las relaciones se harán mucho más evidentes para una mejor comprensión del sistema. Por este motivo se propone un modelo de pronóstico de desarrollo de áreas de conocimiento a través del conocimiento previo generado, o dicho de otra manera, un modelo de pronóstico de desarrollo de disciplinas no desarrolladas previamente por una Universidad, en función de una red de conocimiento generada con anterioridad.

6.2. Descripción del modelo

Para el desarrollo de este capítulo se ha utilizado como modelo la cartografía del conocimiento en Chile (CCC) desarrollada en (Cardenas et al., 2015), el cual es una adaptación del espacio de productos de Hidalgo, et al (Hidalgo et al., 2007), a través de la mirada de la producción científica, utilizando redes de complemento disciplinar.

Es importante señalar que la construcción de la CCC considera las medidas de Actividad de los nodos, así mismo, las características de Centralidad de cada nodo vienen condicionadas por el algoritmo *Maximum Spanning Tree* (Graham y Hell, 1985), que convierte la red de áreas de conocimiento en un árbol (figura 6.1). Estas consideraciones pueden ayudar a comprender el resultado del modelo predictivo que se describe a continuación.

Figura 6.1: Red disciplinar transformada a Cartografía del Conocimiento en Chile.

Es de suponer que en un país desarrollado, toda universidad suficientemente grande y generalista acabará desarrollando la mayoría de las disciplinas científicas, pero no todas con la misma intensidad. Universidades mas pequeñas o especializadas (como las politécnicas), puede que no alcancen a desarrollar áreas muy alejadas de sus intereses. Sin embargo, y parte de esta tesis es un ejemplo claro, cada vez se desarrollan más investigaciones interdisciplinarias que aunan áreas tradicionalmente alejadas, como pudieran ser la física y la sociología (Olivares et al., 2019) y (Morales et al., 2015a). Así, el interés radica en saber qué áreas no desarrolladas lo van a hacer antes.

El utilizar la CCC resume las interacciones que se dan en la cultura investigadora del país entre áreas de conocimiento y refleja el sustrato por el que se relacionarán áreas de conocimiento en cada universidad chilena.

Se dirá que una universidad desarrolla áreas científicas cuando sus investigadores publican en diversas revistas asociadas a esas áreas. Si un investigador de una determinada universidad realiza una publicación, entonces decimos que aquella universidad también ha generado conocimiento en las áreas asociadas a esa revista.

Llamaremos T_i a la cantidad de publicaciones de una universidad en la disciplina i , así se puede proyectar la investigación de una universidad sobre la CCC con ciertas áreas encendidas, $T_i > 0$ (desarrolladas) y otras apagadas $T_i = 0$ (no desarrolladas, pero susceptibles de serlo), como se muestra en las figuras 6.2 y 6.3 para la Pontificia Universidad Católica de Valparaíso. La idea del modelo es proponer una medida que cuantifique la capacidad que tiene una disciplina para ser desarrollada en un futuro próximo, sin importar exactamente el año en que lo hará.

Adicionalmente se definirá el potencial de una disciplina no desarrollada en función de su posición en la red, ya que si áreas vecinas en la CCC están encendidas es de suponer que habrá más posibilidades de ser desarrollada

Figura 6.2: Ejemplo de las disciplinas desarrolladas (de color distinto a gris) y con potencial de desarrollo (color gris) para la Pontificia Universidad Católica de Valparaíso en el período 2008-2014. El tamaño del nodo es proporcional al número de publicaciones por disciplina.

Figura 6.3: Ejemplo de las disciplinas de alto potencial de desarrollo para la Pontificia Universidad Católica de Valparaíso para el año 2015, utilizando el período 2008-2014 con el parámetro $\alpha = 2$.

para saber si la Actividad de los vecinos, es relevante a la hora de pronosticar nuevos desarrollos. Para esto se define b_i como:

$$b_i = \sum_j A_{ij} T_j \quad (6.1)$$

Además, se define la cantidad a_i como el número de vecinos encendidos que tiene una disciplina i no desarrollada. Si A_{ij} es la matriz de adyacencia de la CCC podemos expresar a_i como:

$$a_i = \sum_j A_{ij} d_j \quad (6.2)$$

donde

$$d_j = \begin{cases} 1 & T_j > 0 \\ 0 & T_j = 0 \end{cases} \quad (6.3)$$

En la figura 6.4 se muestra un esquema de proyección de disciplinas sobre la CCC de las áreas desarrolladas por una universidad. Las disciplinas 1, 2 y 6 han sido desarrolladas y las 3, 4 y 5 aún no. En la tabla 6.1 se muestra la cantidad de veces que cada disciplina ha sido desarrollada, b_i es una medida de desarrollo disciplinar ponderado conjuntamente con la Centralidad de la red de la disciplina no desarrollada T_i , y los correspondientes valores de a_i y b_i de las disciplinas con $T_i = 0$.

Figura 6.4: Esquema de proyección de áreas desarrolladas por una universidad (en azul) sobre la CCC.

Disciplina	T_i	a_i	b_i
1	5	—	—
2	3	—	—
3	0	2	8
4	0	0	0
5	0	1	3
6	3	—	—

Tabla 6.1: Valores T_i , a_i y b_i para cada una de las disciplinas de la figura 6.4.

Se dirá que una disciplina tiene alto potencial de desarrollo, es decir, que tiene grandes posibilidades de ser desarrollada en los próximos años, si cumple la siguiente relación:

$$b_i > \alpha a_i \quad (6.4)$$

donde

$$\alpha \geq 0 \quad (6.5)$$

es un parámetro de control que permite valorar la importancia de la actividad de las áreas desarrolladas a la hora de pronosticar el desarrollo de un área vecina. Así, si $\alpha > 1$ se exige que la actividad de alguno de los vecinos sea mayor que la unidad, estableciendo un umbral de actividad de las áreas previamente desarrolladas para predecir nuevos desarrollos. El caso de $\alpha < 1$, teniendo en cuenta que $b > a$ es equivalente a exigir que $b_i > 0$, es decir que la disciplina i tenga al menos una disciplina vecina de la CCC ya desarrollada, independientemente de su actividad. En este caso, el potencial de desarrollo sólo depende de la posición en la red. En la figura 6.3 se muestran las disciplinas de alto potencial de desarrollo con $\alpha = 2$ para la Pontificia Universidad Católica de Valparaíso.

Para validar el modelo y estimar el valor de α , se realizaron los siguientes pasos:

1. Se elige un conjunto de universidades con áreas sin desarrollar.
2. Se obtiene el total de disciplinas que se desarrollaron en esta Universidad, en un determinado intervalo de tiempo junto con el número de veces que cada disciplina apareció en sus publicaciones (T_i).
3. Se proyecta sobre la CCC la estructura de disciplinas encendidas ($T_i > 0$) y apagadas ($T_i = 0$).

4. Se obtiene el total de disciplinas con potencial de desarrollo en función de α , es decir, aquellas que cumplen la relación de la ecuación 6.4.
5. Se comprueba en un periodo posterior cuantas de esas disciplinas han sido realmente desarrolladas.

6.3. Elección de universidades y set de datos

En Chile las universidades pueden clasificarse en dos tipos: las pertenecientes al Consejo de Rectores de Universidades Chilenas (CRUCH), que es el conjunto de todas las universidades estatales y un pequeño número de universidades privadas con función pública, y al resto de Universidades Privadas no adheridas al CRUCH.

Para la validación del modelo hemos considerado a las 20 mejores universidades chilenas según el ranking América Economía del 2018 ((AméricaEconomía, 2018)), la cual incluye a un conjunto de 14 universidades del CRUCH y a 6 Universidades Privadas no adheridas al CRUCH, distribuidas a lo largo de todo el país.

Análogamente a la construcción de las redes de complemento disciplinar, se ha utilizado la base de datos de ISI Thomson Reuters ® para la recopilación de las publicaciones científicas en el período 2008-2018. Este set de datos se ha dividido en dos: el conjunto de datos de disciplinas desarrolladas por una universidad (2008-2014) y otro conjunto de datos para evaluar las predicciones (2015-2018), cuya distribución de publicaciones se puede observar en la tabla 6.2.

Universidad	2008-2014	2015-2018
Pontificia Universidad Católica de Chile	11100	8704
Pontificia Universidad Católica de Valparaíso	2002	2141
Universidad Adolfo Ibáñez	498	801
Universidad Alberto Hurtado	223	259
Universidad Austral de Chile	2609	2039
Universidad Católica del Norte	1714	1349
Universidad de Antofagasta	696	763
Universidad de Chile	13054	10037
Universidad de Concepción	5593	4076
Universidad de la Frontera	1781	1695
Universidad de los Andes	593	762
Universidad de Santiago de Chile	2587	2207
Universidad de Talca	1299	1280
Universidad de Tarapacá	843	820
Universidad de Valparaíso	1600	1549
Universidad del Bío Bío	762	786
Universidad del Desarrollo	549	764
Universidad Diego Portales	1265	1351
Universidad Nacional Andrés Bello	1804	2161
Universidad Técnica Federico Santa María	2398	2167

Tabla 6.2: Distribución de publicaciones para cada una de las universidades seleccionadas para la evaluación de la metodología, separados en conjunto de datos de disciplinas desarrolladas por una universidad (2008-2014) y conjunto de datos de evaluación (2015-2018)

Para evaluar la clasificación realizada es necesario contar con ciertas métricas que indiquen su fiabilidad. Para esto, se utilizarán herramientas que se utilizan para realizar pruebas diagnósticas como son la exactitud y la sensibilidad. Inicialmente se debe desarrollar una tabla de contingencia para el resultado de la prueba, tal como se ve en la tabla 6.4.

	Disciplinas Desarrolladas	Disciplinas no desarrolladas
Positivo	VP	FP
Negativo	FN	VN
Total	VP+FN	FP+VN

Tabla 6.3: Tabla de contingencia.

En donde se tienen los siguientes elementos de prueba a considerar:

- **Verdadero positivo (VP)**: indica la cantidad de veces que las discipli-

nas pronosticadas de la clasificación fue correcta.

- **Verdadero negativo (VN)**: es la cantidad de veces que el resultado de la clasificación indica las disciplinas que no se desarrollarán de manera correcta.
- **Falso positivo (FP)**: indica la cantidad de veces que se indicaba que una disciplina sería desarrollada, pero finalmente no lo hizo.
- **Falso negativo (FN)**: es la cantidad de veces que una disciplina se indicó que no se desarrollaría, cuando realmente fue desarrollada.

Con los datos anteriores es posible armar los métodos diagnósticos para evaluar la clasificación, y estos son:

- **Exactitud**: también llamado fiabilidad, corresponde a la probabilidad de que se obtenga un resultado correcto y se calcula según la ecuación 6.6.

$$exactitud = \frac{VP + VN}{TOTAL} \quad (6.6)$$

donde $TOTAL = VP + VN + FP + FN$

- **Sensibilidad**: esta prueba diagnóstica indica la capacidad del algoritmo para identificar una disciplina correcta y se calcula según la ecuación 6.7.

$$sensibilidad = \frac{VP}{VP + FN} \quad (6.7)$$

Para las pruebas, el parámetro α se ha ido variando para observar en qué punto el pronóstico logra los mejores resultados. Tal como ya se mencionó, se utilizó el período 2008-2014 para obtener las disciplinas desarrolladas por las universidades y el período 2015-2018 como período de evaluación. En la figura 6.5 se muestra la evolución del promedio de la exactitud y de la sensibilidad para los valores de $\alpha = 0$, $\alpha = 0,5$, $\alpha = 1$, $\alpha = 1,5$, $\alpha = 2$, $\alpha = 3$ y $\alpha = 4$.

Para evaluar la bondad de los resultados se ha realizado un modelo nulo, que consiste en, a través de cada una de las disciplinas vecinas desarrolladas en la CCC, seleccionar al azar si esta disciplina será desarrollada o no. Para esto, se realizaron un total de 10 repeticiones.

Tal como se puede observar en la figura 6.5, a medida que el valor de α aumenta, la sensibilidad disminuye, mientras que el valor de la exactitud permanece cercano a la del modelo nulo, de forma que la mayor sensibilidad se obtiene con $\alpha = 0$. Sin embargo la importancia del modelo radica en su sensibilidad, ya que es la forma en que el algoritmo logra identificar correctamente aquellas disciplinas desarrolladas, por lo que se tomará un valor de $\alpha = 0$ como referencia para el siguiente análisis del algoritmo.

Figura 6.5: Evolución del promedio de la exactitud (izquierda) y la sensibilidad (derecha) de las 20 universidades seleccionadas de acuerdo a la evolución del parámetro α .

Un valor de $\alpha = 0$ significa que basta con que se haya desarrollado al menos un vecino de la disciplina no desarrollada, para que ésta sea pronosticada como un futuro candidato de desarrollo, por lo que este algoritmo depende completamente de su Centralidad.

Muchos de los FN tienen que ver con disciplinas que en el período 2008-2014 tenían $b = 0$, pero que al desarrollarse disciplinas vecinas en el período

Figura 6.6: Barra de errores de los promedios de todas las universidades para cada valor de α y del modelo nulo.

de prueba, pasaron a tener un $b > 0$.

Las universidades en las que se obtiene una menor sensibilidad en la predicción son aquellas con menor producción y menos áreas desarrolladas. En estas universidades el desarrollo de nuevas áreas puede estar condicionado por factores externos como puede ser nuevas contrataciones de investigadores, el desarrollo de nuevas titulaciones académicas u otras cuestiones que no se contemplan en la construcción de la CCC. Por otro lado, se observa que, en universidades más grandes, con más áreas desarrolladas, y por tanto con menos por desarrollar, la sensibilidad del modelo con $\alpha = 0$ alcanza el 100

En la tabla 6.4 presentada, se observa que en 9 casos la sensibilidad es de al menos un 0.9 y en 13 de al menos 0.8.

En la figura 6.6 se observa las barras de error de la sensibilidad y la exactitud para cada α . Además se observa que el modelo nulo tiende a 0,5 tanto en exactitud y sensibilidad. De hecho el promedio de ambos es de 0,497 ($\sigma=0.22$) y 0,506 ($\sigma=0.04$) respectivamente, mientras que el promedio para el algoritmo con $\alpha = 0$ es de 0,5 ($\alpha = 0,09$) y 0,857 ($\alpha = 0,13$) respectivamente.

De los resultados anteriores se puede observar que, para la exactitud, la metodología propuesta no difiere mucho del modelo nulo, sin embargo la sensibilidad si representa una gran diferencia y es efectivamente lo que se busca potenciar, ya que esta metodología pronostica que disciplinas serán desarrolladas en el futuro, por lo que sigue existiendo margen de tiempo para que aquellas disciplinas sean desarrolladas y que la exactitud mejore.

Por otro lado, se observa que la Actividad no importa en el modelo, ya que cuando se potencia este concepto con valores de $\alpha > 0$, se puede observar una drástica disminución de la sensibilidad a medida que crece, en donde con $\alpha = 4$ se obtienen valores que son prácticamente equivalentes al modelo nulo.

Finalmente se considera que la CCC es una herramienta útil que puede permitir a las instituciones potenciar ciertas áreas del conocimiento no generadas enfocándose en aquellas que ya han sido desarrolladas, es decir, si se desea desarrollar una disciplina B y sabemos que la disciplina A ha sido desarrollada por aquella institución y ambas se encuentran relacionadas dentro de la cartografía del conocimiento, entonces la idea es enfocar los esfuerzos en la gente que desarrolla la disciplina A para poder expandir el desarrollo de la nueva disciplina B en un futuro cercano.

Universidad	Desarrolladas (2008- 2014)	Nuevas (2015- 2018)	VP	FP	FN	VN	Exactitud	Sensibilidad
Universidad de Chile	238	7	4	3	0	0	0.57	1
Pontificia Universidad Católica de Chile	238	6	3	3	0	0	0.5	1
Universidad de Concepción	212	13	11	15	0	1	0.44	1
Pontificia Universidad Católica de Valparaíso	180	32	20	23	9	7	0.46	0.69
Universidad de Santiago de Chile	205	20	14	14	5	1	0.44	0.74
Universidad Austral de Chile	198	14	12	23	2	4	0.39	0.86
Universidad de Talca	166	34	30	35	3	3	0.46	0.91
Universidad Técnica Federico Santa María	140	26	16	33	9	39	0.57	0.64
Universidad de los Andes	144	34	30	44	2	18	0.51	0.94
Universidad Diego Portales	192	16	12	31	3	4	0.32	0.8
Universidad de la Frontera	185	23	20	31	1	2	0.41	0.95
Universidad Andrés Bello	174	33	27	31	3	5	0.48	0.9
Universidad de Valparaíso	193	20	18	27	1	2	0.42	0.95
Universidad Adolfo Ibañez	131	48	30	27	13	36	0.62	0.7
Universidad Católica del Norte	187	26	22	29	1	2	0.44	0.96
Universidad del Desarrollo	136	51	37	33	12	20	0.56	0.76
Universidad de Tarapacá	158	24	19	50	3	8	0.34	0.86
Universidad del Bío-Bío	135	35	27	44	6	25	0.51	0.82
Universidad Alberto Hurtado	60	35	16	32	16	115	0.73	0.5
Universidad de Antofagasta	125	22	18	53	3	38	0.5	0.86

Tabla 6.4: Distribución de disciplinas desarrolladas en el período 2008-2014 y 2015-2018, en conjunto con la matriz de confusión y los valores de pruebas diagnósticas exactitud y sensibilidad para cada una de las 20 universidades seleccionadas.

Capítulo 7

Herramienta de visualización de redes de colaboración científica

En este capítulo se presentará una herramienta de visualización de redes de colaboración científica para autores chilenos y de complemento disciplinar, llamada RedSearch (<http://www.redsearch.cl>) y cuya metodología de desarrollo está basada en los resultados de los capítulos previos. Esta herramienta consiste en un buscador web de investigadores, instituciones y disciplinas que muestra sus relaciones y principales actores del sistema a través de una sencilla visualización de grafos. La interfaz gráfica del buscador se observa en la figura 7.1

Figura 7.1: Interfaz de búsqueda de RedSearch.

7.1. Método de construcción

RedSearch se construyó para ser utilizado a través de cualquier navegador web, usando los lenguajes de programación PHP y Python y JavaScript y el lenguaje de etiquetados HTML5. La base de datos utilizada corresponde a la proporcionada por ISI Thomson Reuters a través de su API, utilizando la condición de que al menos uno de los autores de cada publicación indexada tenga alguna afiliación chilena. De esta forma se generan tres tipos de redes a las que se le denominó Redes de Personas, Redes de Instituciones y Redes Disciplinarias. Adicionalmente existe una herramienta llamada Conecta, que entrega un grafo bipartito de conexión mínima entre dos usuarios cualesquiera.

A continuación se describirán las características de estas redes.

7.2. Redes de Personas

Para el caso de las redes de personas, la estructura de generación de los datos funciona de acuerdo a la estructura de la figura 7.2. Se obtienen las

Figura 7.2: Esquema de datos y listado de autores para la Publicación 1.

publicaciones desde la base de datos de WoS y de cada publicación se obtiene un listado de autores. Todos los autores de esta publicación se enlazan entre ellos y el resultado final es un grafo con múltiples componentes conec-

tadas con todos los autores asociados a Chile. Para nuestro caso usaremos el siguiente ejemplo, realizaremos una búsqueda de relaciones de los autores Eric Goles Chacc, Juan Pablo Cárdenas y Gastón Olivares Fernández. Con esto obtenemos el grafo de la figura 7.3. En esta red se pueden observar dos

Figura 7.3: Red de autores para Eric Goles, Juan Pablo Cárdenas y Gastón Olivares. El color del nodo es de acuerdo a la afiliación asociada, su tamaño es proporcional al número de publicaciones y el peso del enlace corresponde al número de interacciones entre dos autores.

componentes conexas con todas las coautorías de los usuarios.

7.3. Redes de Instituciones

Análogamente, para el caso de las redes de instituciones se realiza el mismo procedimiento de obtención de datos que para las redes de autores, con la diferencia que para este caso vinculamos las instituciones asociadas a cada una de las publicaciones, dando como resultado final un grafo con

múltiples componentes conectadas de instituciones chilenas y extranjeras que han colaborado con Chile. El esquema de extracción de datos se puede ejemplificar en la figura 7.4. Siguiendo el ejemplo de los autores buscados en

Figura 7.4: Esquema de datos para una red de instituciones.

las redes de autores previamente mencionadas, también se puede obtener la red de instituciones, como se observa en la figura 7.5. Esta red muestra las relaciones que existen entre todas las instituciones de la búsqueda realizada, a través de las afiliaciones de los autores de la publicación.

7.4. Redes de Disciplinas

Finalmente se tienen las redes de disciplina (llamadas red de complemento disciplinar en los capítulos previos), cuyo método de construcción es análogo a los dos anteriores. Se extrae las disciplinas científicas asociadas a la revista de una publicación y se generan vínculos entre todos ellos, tal como se describió en el capítulo 5. Luego se obtiene el grafo total con múltiples componentes conectados de disciplinas de Chile y el esquema de construcción se observa en la figura 7.6.

Nuevamente utilizando el ejemplo de las dos redes anteriores, obtenemos la red de disciplinas de acuerdo a las disciplinas científicas asociadas a las publicaciones de los autores seleccionados, lo que se aprecia en la figura 7.7. En ella cada color representa una área de la OECD ya descrita en esta tesis

Figura 7.5: Red de Instituciones de acuerdo a la afiliación de los autores que surgen a partir de la búsqueda de Eric Goles, Juan Pablo Cárdenas y Gastón Olivares.

y que se puede observar en mayor profundidad en el Apéndice C, en este caso en naranja se observan las disciplinas asociadas a Ciencias Básicas, en azul a Medicina, en verde a Ciencias de la Ingeniería y finalmente en rojo a Humanidades.

7.5. Red Conecta

Esta herramienta consiste en la búsqueda de un camino mínimo entre dos autores cualesquiera afiliados a alguna institución chilena. La estructura de este tipo de grafo se puede observar en la figura 7.8. La idea principal consiste en generar algún vínculo entre un autor A y un autor B cualquiera. Si se considera el supuesto de que las redes de científicos suelen ser pequeñas, obteniendo de alguna forma propiedades de redes de mundo pequeño, podemos decir que a un pequeño número de pasos podemos llegar desde A hasta B atravesando un pequeño número de autores.

Sin embargo el problema es que a esta escala, el grafo de autores tiene muchas componentes conexas que hacen imposible la generación de un camino entre estos dos autores.

Figura 7.6: Esquema de datos para red de disciplinas.

Para solucionar lo anterior, se añadieron las afiliaciones institucionales de un usuario como nueva componente a la red, obteniendo finalmente una especie de red bipartita de autores e instituciones. Por ejemplo, si un autor A está relacionado con una institución I y además un autor B está relacionado con la misma institución I, entonces se genera un vínculo entre el autor A, la institución I y el autor B.

De esta forma, en el caso de que no haya un camino más corto entre dos autores y considerando el ejemplo anterior, las instituciones sirven de intermediarios para poder llegar desde A a B.

Para ejemplificar mejor, si se quisiera ver la distancia mínima de pasos entre Gastón Olivares Fernández y Eric Goles Chacc, la herramienta desplegaría los datos de la figura 7.9. En la figura 7.9 se puede observar a los nodos seleccionados representados con un cuadrado, mientras que los otros autores son círculos y las instituciones son rombos.

En este ejemplo se observa que ambos autores se encuentran en componentes conexas distintas de la gran red de autores, sin embargo las instituciones ayudan a generar un vínculo, permitiendo que ambos autores tengan 4 grados de separación.

Figura 7.7: Red de Disciplinas de acuerdo a las disciplinas asociadas a las publicaciones de Eric Goles, Juan Pablo Cárdenas y Gastón Olivares.

7.6. Ranking de Autores

El ranking de autores es generado cuando se realizan búsquedas de redes de coautoría a través instituciones o disciplinas y su ecuación está basada de acuerdo a los datos obtenidos en el capítulo 4.

De esta forma, el ranking está definido por la ecuación 7.1 aplicada a cada autor y ordenada de mayor a menor.

$$Ranking(A) = Cen(A) * ef_A \quad (7.1)$$

donde A es un autor cualquiera que se encuentre dentro de los parámetros buscados. Además la Centralidad está definida por la ecuación 7.2.

$$Cen(A) = PR(A) * CC(A) * k_A \quad (7.2)$$

donde $PR(A)$ es el Pagerank del nodo A , $CC(A)$ es el número de nodos de la componente conexa en la cual se encuentra el autor A y k_A es el grado de A . Por otro lado la eficiencia ef_A es definida como:

$$ef_A = \frac{cit_A}{pub_A} \quad (7.3)$$

Figura 7.8: Estructura básica de un grafo Conecta. En la figura se observan dos autores y una institución, en donde ambos autores están conectados entre sí debido a una coautoría en conjunto y además por estar afiliados en la misma institución.

donde cit_A es el total de citas totales del autor A dentro de la red y pub_A es el total de publicaciones dentro de la red del autor A .

En la figura 7.10 se puede observar el ranking de autores desplegado por la herramienta para la red de Biología en el año 2017.

LIDERAZGO EN LA RED	
1.	BACIGALUPE-CIRILLO, LEONARDO DANIEL
2.	BOZINOVIC-KUSCEVIC, FRANCISCO...
3.	TORRES-SAAVEDRA, RODRIGO FERNANDO
4.	GAITAN-ESPITIA, JUAN DIEGO
5.	NESPOLO-ROSSI, ROBERTO FERNANDO
6.	GIANOLI-MOLLA, ERNESTO MARIO
7.	LUCERO, CLAUDIA
8.	DIAZ, HUGO
9.	TOLEDO-RIVERA, CAMILO ENRIQUE
10.	DEL RIO-TRONCOSO, RODRIGO ANDRES
11.	BARTHELD-VILLAGRA, JOSE LUIS
12.	ITURRIAGA, RODRIGO
13.	MARQUET, PABLO A.
14.	SANTIVIAGO, CARLOS A.
15.	MALIQUEO, MANUEL

Figura 7.10: Ranking de autores para la red de la disciplina Biología en el año 2017 para Chile.

7.7. Datos de un autor

Dentro de las opciones, también se puede observar en profundidad los datos meritocráticos y topográficos de cada autor. Si utilizamos como ejemplo al líder de la red anterior de Biología, Leonardo Bacigalupe Cirillo, el primer resumen que se obtiene se aprecia en la figura 7.11. En la figura 7.11 se pueden observar las disciplinas en las que ha trabajado el autor, sus afiliaciones institucionales y finalmente sus publicaciones y citas totales.

Por el lado de sus estadísticas se pueden observar además cual es su preferencia de conectividad, es decir, si sus coautores son parecidos a él en términos de grado, citas o publicaciones, tal como se aprecia en la figura 7.12.

Para este caso se observa que no tiene preferencia de conexión y además se obtienen los coeficientes de correlación de Pearson para las publicaciones, citas y grado y la mediana (en azul su valor) de la comunidad en la que se encuentra (sus coautores), los que se muestran en la figura 7.13.

Finalmente también es posible la distribución y comparativa de disciplinas que ha desarrollado el autor frente a las disciplinas que ha desarrollado su comunidad. En este caso se puede observar que las disciplinas que desarrolla la comunidad de este usuario es bastante más extensa a las que él desarrolla, por lo que este autor tiene un potencial de crecimiento en cuanto a desarrollo de disciplinas si se apoya mucho más en sus coautores.

Figura 7.9: Ejemplo de camino mas corto para los autores Eric Goles y Gastón Olivares.

Figura 7.11: Resumen del autor Leonardo Bacigalupe Cirillo.

Figura 7.12: Preferencia de conexión del autor Leonardo Bacigalupe Cirillo.

Figura 7.13: Comparativa de las disciplinas desarrolladas por Leonardo Bacigalupe Cirillo y sus coautores.

Capítulo 8

Redes sociales virtuales: Polarización en Twitter

Este capítulo es una expansión de los estudios previos realizados en redes sociales presenciales, utilizando redes sociales *online*. Para ello, se utilizará la plataforma social Twitter para analizar la polarización política surgida durante la semana previa a las elecciones electorales chilenas del 2017, utilizando redes de *Retweet*

8.1. Introducción

Una sociedad perfectamente polarizada es aquella que está dividida en dos grupos de un mismo tamaño, con la totalidad de las opiniones confrontadas. De acuerdo a esta definición, los autores de (Morales et al., 2015c) proponen una metodología para estudiar y medir la emergencia de la polarización de interacciones sociales. Sin embargo, la polarización social puede emerger de diversos escenarios, como por ejemplo, durante campañas electorales. Mas aún, los procesos electorales en donde hay sólo dos visiones políticas incrementan la posibilidad de encontrar una alta polarización política (Sundquist, 1983). Por ejemplo, los ciudadanos de los Estados Unidos de América se encuentran frente a un escenario polarizado durante las elecciones presidenciales, donde ellos tienen que votar de acuerdo a decisiones dicotómicas entre dos partidos: republicanos y demócratas. Estudios previos han mostrado que las elecciones de los Estados Unidos de América polarizan la Internet, donde tanto Twitter (Conover et al., 2011), (Wang et al., 2017) y blogs (Adamic y Glance, 2005) manifiestan una alta polarización política.

Otros escenarios políticos que requieren decisiones dicotómicas son aquellos procesos electorales que requieren una segunda vuelta. Mientras que en la primera vuelta los ciudadanos pueden votar en un amplio espectro de partidos políticos, en la ronda final ellos deben votar finalmente sólo por dos

candidatos. En efecto, algunos votantes podrían no sentirse completamente identificados por algún sector, pero aún así deben decidirse por un lado. Estudios previos muestran que en las segundas rondas electorales se incrementa la polarización política en los países (Serra, 2018).

Durante los últimos años, plataformas de redes sociales como Twitter, se han convertido en medios efectivos para que los partidos políticos puedan difundir sus mensajes y campañas para sus candidatos, llegando a una gran cantidad de la población a muy bajo costo (Huberman et al., 2009), (Herdagdelen et al., 2013) y (Parmelee, 2013). A través de los datos de Twitter que se encuentran disponibles públicamente por medio de su API, científicos pueden analizar estas campañas electorales caracterizando estrategias políticas, analizando la influencia de los usuarios (Solop, 2010), (Morales et al., 2012) y (Conover et al., 2012) y modelar el comportamiento humano en estos procesos (Gonçalves et al., 2011), (Borondo et al., 2012), (Jungherr, 2013), (Borondo et al., 2015) y (Yaqub et al., 2017).

En este capítulo se presenta un estudio que analiza la última semana de la segunda vuelta de las elecciones presidenciales chilenas, en donde los votantes debieron elegir entre los candidatos Sebastián Piñera, Alejandro Guillier o la abstención. Nos enfocamos en el análisis de la polarización política que emerge y en realizar un seguimiento de su evolución durante la semana anterior a las elecciones y en el último día de la votación. Para eso, se estimó la opinión de los usuarios de Twitter a través de un pequeño conjunto de usuarios influyentes, a los que se llamó usuarios de la élite, cuya opinión es conocida. Posteriormente, se han medido los resultados políticos a través de polarización y se ha analizado su evolución durante la última semana del día de la votación. Se ha encontrado un cambio en las opiniones de los usuarios y en la polarización política durante el día de la votación. Además, se exploró hasta que punto este cambio en el comportamiento se explica por la participación de nuevos usuarios que comentan sobre las elecciones ese mismo día, o porque los usuarios cambiaron de opinión durante el último día. Finalmente, se muestra que el incremento de la polarización es explicado por un comportamiento de propaganda hacia los candidatos de los usuarios que ya estaban comprometidos con la conversación. Sin embargo, se ha demostrado que la disminución de la polarización observada el día de la elección fue causada por nuevos usuarios que no estaban tan comprometidos con el debate político, los que entraron a la conversación actuando como puentes entre ambas partes polarizadas.

8.2. Sistema Electoral

8.2.1. Sistema electoral chileno

El sistema de elecciones presidenciales es realizado cada cuatro años. Cada chileno que cumple los requerimientos que el servicio electoral chileno establece, puede presentar a un candidato para ser elegido en las próximas elecciones presidenciales de la República de Chile. Un candidato puede pertenecer a un partido político o a una lista de partidos, bastando que esta comunidad inscriba a su nombre como candidato. Si no existiera un consenso dentro de una comunidad política respecto a qué candidato postular, es posible realizar unas primarias para definir un único candidato. En el caso de los candidatos independientes, estos tienen la obligación de no estar afiliados a ningún partido político, al menos dos meses antes del registro de sus candidaturas. Ellos tienen la necesidad de tener al menos un 0.5 % de firmas del número total de votantes de las últimas elecciones disputadas, siendo 33493 firmas al año 2017. En el día de la elección, cada ciudadano calificado puede votar por un sólo candidato. Cuando un candidato obtiene una mayoría absoluta ($>50\%$), este candidato será presentado como «Presidente de la República de Chile». En caso contrario, se procede a realizar una segunda ronda presidencial, en donde los dos candidatos con mayor número de votos pasan a un nuevo proceso electoral. En las elecciones del 2017, los partidos de centro izquierda no lograron ponerse de acuerdo y varios candidatos fueron presentados, denotando una clara división dentro de los partidos políticos. Esto no fue el caso de los partidos de centro derecha, quienes postularon a un único candidato.

Figura 8.1: Resultados de la primera ronda de las elecciones presidenciales chilenas del 2017.

En la primera ronda de las elecciones del 2017 realizadas el 19 de noviembre, tuvo un ratio de abstención del 53.3% y no se logró obtener una mayoría absoluta para ningún candidato (ver figura 8.1).

La segunda ronda fue realizada un mes después, el 17 de diciembre, con una tasa de participación ligeramente mayor (51%). Los resultados dieron por ganador de las elecciones a Sebastián Piñera, candidato de la centro derecha (ver figura 8.2).

Figura 8.2: Resultados de la segunda ronda de las elecciones presidenciales del 2017.

8.2.2. Descripción de los datos

Para construir el conjunto de datos analizado para este estudio, se han descargado todos los mensajes de Twitter relacionados a las elecciones presidenciales chilenas del 2017 durante la última semana de la campaña de la segunda ronda electoral (semana que incluye el día de la votación). Todos los *tweets* fueron obtenidos usando la API de Twitter. Esta API permite la descarga de tweets de acuerdo a un conjunto de palabras claves. Para construir el conjunto de datos hemos elegido las siguientes palabras claves: #chileelige, #eleccionesenchile, #elecciones2018, #chilevota, #elecciones2017, #eleccioneschile2017, #pinera y #guillier.

El conjunto final de datos fue compuesto por 203.612 generados por un total de 68.048 usuarios distintos entre el 11 y 17 de diciembre del 2017. El 83.5 % del total de *tweets* correspondía a *retweet*.

8.3. Metodología

8.3.1. Redes de retweet

Se pueden considerar los *retweets* como un agente de influencia entre los usuarios. Por lo tanto, cuando el usuario A reenvía un *tweet* publicado

originalmente por el usuario B, se puede inferir que B está influyendo en A (Borondo et al., 2015), (Boyd et al., 2010) y (Morales et al., 2015a). En las redes de *retweet*, cada nodo corresponde a un usuario, y se establece un enlace desde el nodo A al nodo B cuando el usuario A reenvía un mensaje publicado por el usuario B. Este enlace se dirige con la dirección de A a B. Por lo tanto, el grado de entrada en un nodo dado mide el número total de usuarios diferentes que lo han retuiteado, y el peso de entrada, el número de *retweets* totales que recibió.

En la figura 8.3(a), se muestra la red de retweet correspondiente para el período de la campaña electoral, excluyendo el día de la votación (11-16 de diciembre). La red está compuesta por 26.175 nodos y 48.141 enlaces. La red correspondiente a la campaña electoral completa, incluido el día de la votación, se muestra en la figura 8.3 (b) para la comparación. Esta red tiene un total de 68.048 nodos y 170.045 enlaces. En la figura, el tamaño de los nodos es proporcional a su grado.

La distribución en grados de ambas redes es muy heterogénea con un pequeño número de usuarios que acumulan un alto porcentaje de los *retweets* (ver figura 8.4). La mayoría de los usuarios (57 %) solo *retwitean* un solo *tweet* durante el período completo analizado, mientras que una fracción muy pequeña de usuarios (0.025 %) *retwitteó* más de 100. Se observa una tendencia similar al analizar la distribución de grados de entrada. La mayoría de los usuarios no reciben ningún *retweet*; y observamos que el 83.5 % de los usuarios tienen $k_{in} = 0$ y el 7 % de los usuarios tienen $k_{in} = 1$. En contraste, estos usuarios conviven junto con una minoría de usuarios extremadamente influyentes que recibieron muchos *retweets*. De hecho, el 0.075 % de los usuarios recibió más de 500 *retweets* e incluso encontramos un usuario cuyos *tweets* recibieron más de 5000 retweets. Estos usuarios altamente *retwiteados* son los verdaderos influyentes que gobiernan la conversación.

Figura 8.3: Redes de retweets de la segunda ronda de las elecciones presidenciales chilenas del 2017, correspondiente a la campaña electoral del período del 11 al 16 de diciembre (a) y del 11 al 17 de diciembre (b). El tamaño de los nodos es proporcional a su grado de entrada k_{in} y sus colores están asociados a la opinión de acuerdo a la escala que se muestra en la leyenda: -1 Guillier, 0 Neutral y 1 Piñera.

Figura 8.4: Distribución del grado de entrada y de salida correspondiente a la red de *retweet* para el período del 11 al 17 de diciembre de 2017.

Para analizar adecuadamente esta conversación, se consideraron varios períodos de tiempo. Así, se construyó una red de *retweet* para cada día D , donde la ventana de tiempo considerada para la red de cada día D es la acumulada desde el día 0 hasta el día D .

8.3.2. Inferiendo la opinión de los usuarios

Para inferir la opinión de los usuarios que participan en la conversación de Twitter y medir la polarización política resultante, se utilizó la metodología introducida en (Morales et al., 2015c). En ese trabajo, se presenta un modelo para estimar las opiniones de los usuarios que interactúan en una red social, a través una pequeña cantidad de usuarios cuya opinión es conocida. En el modelo se tienen dos tipos de usuarios, *élite* y *oyentes*. El modelo asume que se conoce con certeza la opinión de la *élite*. Por lo tanto, los usuarios de *élite* tienen una opinión fija que permanece constante y actúa como semillas de influencia. En contraste, la opinión de los *oyentes* es desconocida y se estimará a partir de sus interacciones sociales. El modelo se inicializa asignando un valor de opinión X de -1 o 1 a cada *élite* y un valor inicial $X = 0$ a cada *oyente*. La opinión de cada *oyente* se actualiza iterativamente como

el valor de opinión medio de sus vecinos salientes. Así, la opinión en el paso de tiempo, t , de un oyente dado i , viene dada por la siguiente expresión:

$$X_i(t+1) = \frac{\sum_j A_{ij} X_j(t)}{k_i^{in}} \quad (8.1)$$

donde A_{ij} representa los elementos de la matriz de adyacencia de la red, que toma valor 1 si existe un enlace entre j e i , y k_i^{in} corresponde a su grado de salida. El proceso se repite hasta que todos los nodos convergen a su valor X_i respectivo, en el rango $-1 \leq X \leq 1$. Por lo tanto, los resultados del modelo se dan en una distribución de densidad de los valores de opinión de los nodos $P(X)$. En este caso, la red utilizada es la red de *retweet*. Se genera una red para cada día (D) del periodo analizado; esta red es la red de *retweet* acumulada desde el día 0 hasta el día D . La matriz de adyacencia correspondiente para cada día D viene dada por la siguiente fórmula:

$$A_{ij}(D) = \begin{cases} 1 & \text{si } i \text{ retuitea a } j \in t \leq D \\ 0 & \text{en otro caso} \end{cases} \quad (8.2)$$

Para cada día se ha obtenido una medida de polarización política de la conversación desde la resultante distribución de densidad de la opinión de los nodos $P(X)$. Esta polarización se obtiene como sigue:

$$\rho = (1 - |\Delta A|)d \quad (8.3)$$

donde

$$|\Delta A| = |P(X > 0) - P(X < 0)| \quad (8.4)$$

y $2d$ es la distancia entre el promedio positivo y negativo de las opiniones. Esta fórmula entrega un $\rho = 1$ cuando la distribución está perfectamente polarizada. En este caso la función de la distribución de opinión es dos delta Dirac centrado en -1 y $+1$ respectivamente. Por otro lado, $\rho = 0$ significa que las opiniones no están polarizadas en absoluto.

8.3.3. Usuarios de Élite

Para seleccionar a los usuarios de élite que serán utilizados para inferir la opinión del resto de los usuarios, lo primero que hacemos es seleccionar a los nodos centrales de acuerdo a su topocracia, es decir, su grado de entrada $k_{in} > 200$. Luego de aplicar este filtro se obtienen 159 nodos (ver figura 8.5). De ellos se han descartado 44 usuarios debido a que son cuentas institucionales, medios de información masiva o cuentas que han sido eliminadas por Twitter durante la campaña.

Figura 8.5: Subred de la red de *retweets* correspondiente a la conversación de Twitter de las elecciones chilenas del período 11 al 17 de diciembre de 2017, correspondientes a los 160 nodos con $k_{in} > 200$. Los nodos de la élite asociados a Piñera en rojo y los asociados a Guillier en azul. El resto de los nodos en gris no están incluidas en las élites. El tamaño de los nodos corresponde a su respectivo k_{in} .

Se ha asignado la opinión de +1 o -1 a aquellos nodos influyentes evaluando sus opiniones al analizar sus perfiles de usuario y los *tweets* de cada uno de ellos. Al realizarlo se han identificado 12 cuentas que pertenecerían a políticos (6 de cada partido político). A estas cuentas se les asignó un valor de polarización de acuerdo a su preferencia política. Para el resto de las cuentas, 31 de ellas expresaron su opinión política en su perfil de Twitter (17 en favor a Piñera y 14 apoya a Guillier). Luego, a través de técnicas de minería de datos descritas en el Apéndice A, se han identificado a un grupo de usuarios expresando su opinión de voto por cada candidato (16 de ellos para Piñera y 22 para Guillier). Para asignar el valor a los restantes 34 usuarios, se ha entrenado algoritmos de aprendizaje automático utilizando los *tweets* de cada usuario clasificado. Para esto se ha entrenado una Máquina de Soporte Vectorial (SVM) con kernel lineal (Srivastava y Bhambhu, 2010). Se ha validado el entrenamiento a través de validación cruzada *k-fold*, obteniendo una

precisión promedio del 82 %. Finalmente, se ha usado el modelo entrenado para clasificar los restante 34 usuarios. Un total de 64 fueron clasificados en favor de Guillier, a quienes se les asignó un valor de -1 y 51 usuarios en favor de Piñera, a quienes se les asignó un valor de +1.

Una vez que la élite fue seleccionada, la opinión de los siguientes nodos es generada a partir de la ecuación 8.1). La figura 8.6 muestra la rápida convergencia de este proceso de 100 nodos aleatorios.

Figura 8.6: Convergencia de las opiniones inferidas a través de un proceso iterativo correspondiente al modelo de formación de opinión para 100 nodos seleccionados aleatoriamente. Se debe notar que muchos nodos obtienen las mismas dinámicas de convergencia.

8.4. Resultados

Desde las distribuciones de opinión obtenidas con el modelo para cada uno de los días para el conjunto de datos, se puede observar la evolución de las opiniones de los usuarios que participaron en la conversación de la campaña electoral chilena del 2017. Así, se han usado estas distribuciones para analizar las opiniones de los usuarios y la evolución de la polarización política a través de la campaña electoral. Se ha mostrado que durante todo el período analizado se observan dos picos en la distribución para las dos opiniones confrontadas con -1 y +1. Además, se han identificado diferencias significativas en la distribución entre el día del voto (D 17) y la semana previa. Para explorar estas diferencias se han visualizado ambas distribuciones de opinión en los paneles superiores de la figura 8.7.

En esta figura se ha comparado la distribución del día previo al voto (D 16) y el día del voto (D 17). Las distribuciones muestran que el día previo al de las elecciones, la mayoría de los usuarios tiene un valor de opinión $X = +1$ (opiniones completamente en favor del candidato Piñera). Sin embargo, se ha observado un cambio en este comportamiento durante el día de la votación,

en donde la mayoría de los usuarios tienen una opinión $X = -1$. En el día de la votación se ha encontrado un incremento en las opiniones moderadas a través del candidato Guillier, con un significativo incremento de los usuarios con $P(-1 < X < 0)$. Así, los resultados muestran un claro cambio en las opiniones expresadas por los usuarios de Twitter entre el día previo a las elecciones y el día de las votaciones. Mientras la opinión expresada en Twitter se mantuvo aproximadamente constante durante toda la semana previa al día de las elecciones, esto no fue el caso para el día de la votación.

A pesar de este cambio en las opiniones, no se puede afirmar que los usuarios cambiaron individualmente su opinión durante el día de la votación, ya que la participación generalmente aumenta durante el día de la elección para atraer nuevos usuarios. Así, para aclarar si la distribución obtenida el día de la votación se debe a un cambio significativo en la opinión de los usuarios que participaron activamente durante la semana anterior o a la participación de nuevos usuarios, se analizó por separado la opinión de ambos grupos. En el panel inferior de la figura 8.7, se muestra la distribución de opiniones resultante para el día de la votación, incluidos solo los usuarios que ya participaron durante los días anteriores.

Figura 8.7: Paneles superiores: distribución de opinión estimada para las redes de *retweets* correspondiente a la campaña electoral del período 11 al 16 de diciembre del 2017 (izquierda) y a la campaña electoral del período 11 al 17 de diciembre 2017 (derecha). Panel inferior: distribución de opinión de las redes de *retweets* correspondiente a la campaña electoral del período 11 al 17 de diciembre de 2017 eliminando a aquellos usuarios que sólo participaron durante el día de la votación.

La distribución de opiniones de estos usuarios difiere significativamente de la trazada en los paneles superiores que incluye a todos los usuarios. La población de usuarios que participaron durante la semana anterior aún mantiene una mayoría de valores extremos en torno a 1, destacando que la mayoría de las opiniones en torno a -1 provienen de los nuevos usuarios que se unieron a la conversación el último día de votación. Sin embargo, una fracción de los usuarios que ya participaron durante la semana de la campaña variaron ligeramente su opinión ya que hay un aumento en las opiniones negativas moderadas para ambas poblaciones.

Para comprender mejor la evolución de las opiniones durante el período analizado, se han visualizado las series de tiempo para las opiniones negativas y positivas en la figura 8.8. En esta figura, los triángulos representan opiniones a favor del candidato Piñera y los círculos opiniones a favor del candidato Guillier.

Figura 8.8: Evolución de la distribución de probabilidad de las opiniones para los valores X : -1, +1, >0 y <0 (arriba) y los índices de polarización política, ρ , para las redes de *retweet* comenzando el 11 de diciembre hasta el día D de las elecciones presidenciales chilenas de 2017. Los valores de la línea punteada corresponde a los de la campaña electoral del período 11 al 17 de diciembre de 2017 sin tomar a aquellos usuarios que sólo participaron el día de la votación.

La figura 8.8 refleja que las opiniones se mantuvieron constantes durante los días anteriores al día de la votación, con una mayoría de opiniones a favor de Piñera. Esta tendencia cambió durante el día de la votación, donde más usuarios participaron en la conversación y la fracción de opiniones moderadas a favor de Guillier superó a las de Piñera. Sin embargo, el aumento significativo de opiniones moderadas no se explica por los nuevos usuarios, sino por el cambio de mentalidad durante el día 17 del resto de los usuarios. Si se limita el análisis a las opiniones radicales (+1 o -1), se observa este mismo patrón, con una transición de una mayoría de +1 a una mayoría de -1. Sin embargo, este cambio en las opiniones radicales se debe a los nuevos usuarios que comentan por primera vez durante el último día.

A continuación, se analiza cómo evolucionó la polarización política durante la semana de la campaña electoral. Con este fin, se calculó el índice de polarización política para cada día y se visualizaron las series de tiempo en el panel inferior de la figura 8.8. Se encontró que la conversación se polarizó durante todo el período electoral, con valores en torno al 60 % que aumentaron hasta el 70 % justo antes del día de la votación. Este hecho manifiesta que durante el día anterior a la votación la conversación alcanzó su máxima confrontación. En contraste, el día de la votación, se encontró una disminución abrupta en la que el índice de polarización alcanzó un valor mínimo de 40 %. Sin embargo, si no se consideran los nuevos usuarios (que solo participaron el día 17) y se limita el análisis de polarización política a los usuarios que participaron durante la semana completa, también se observa esta transición, pero la disminución de la polarización política en el día de la votación no es tan brusca y recupera su valor típico. Por lo tanto, el valor mínimo extremo observado el día de la votación se explica por la aparición de nuevos usuarios que participaron en la conversación este día y por la disminución de opiniones extremas que dieron lugar a opiniones más moderadas que se concentraron en el lado negativo (ver figura 8.7). Por lo tanto, a la luz de estos resultados, se puede concluir que los nuevos usuarios que se unieron a la conversación el día de la votación, provocaron una variación significativa de la opinión política general, con un aumento en el flujo de mensajes entre las dos partes.

Para comprender mejor la evolución de la polarización política durante la última semana de las elecciones, se analizó el perfil de los usuarios y los *tweets* publicados durante la semana. Más particularmente, se analizó el uso de menciones en *tweets*. Las menciones son un mecanismo de interacción de Twitter que generalmente se usa para dirigir *tweets* a usuarios específicos, para referirlos o para intercambiar información. En el panel superior de la figura 8.9, se muestran las series temporales de la fracción de *tweets* que contenía al menos una mención. El porcentaje de *tweets* que contienen menciones se mantuvo constante durante la semana con un valor de aproximadamente el 80 %. Sin embargo, para el día de la votación, este valor

disminuyó hasta alcanzar un mínimo de 76 %. La disminución en el número de menciones durante el día de la votación muestra que durante ese día el debate amplía su alcance, no dirigiendo tantos *tweets* a usuarios específicos.

Figura 8.9: Evolución en el porcentaje de menciones (arriba) y el índice de heterogeneidad (abajo), para las redes de *retweet* comenzando el 11 de diciembre hasta el día D de las elecciones presidenciales chilenas del 2017. En el panel superior, la línea punteada corresponde a la campaña electoral del período 11 al 17 de diciembre de 2017 sin considerar aquellas cuentas de usuarios que sólo participaron el día de la votación.

Los *tweets* sin menciones no están dirigidos a nadie en particular y pueden usarse para expresar sentimientos. Una reducción en el número de menciones generalmente implica una transición de racional a más emocional, lo que puede explicar la disminución en el índice de polarización política descrito

anteriormente.

A continuación, se exploraron las razones que explican la disminución en el porcentaje de *tweets* que contienen menciones para revelar si se debe a los nuevos usuarios que se unieron a la conversación el día de la votación o no. Se descubrió que, de hecho, esto se explica por los nuevos usuarios que participaron en el debate el último día, además la fracción de menciones por *tweets* permanece constante durante todo el período. A continuación, se analizó la diversidad de los usuarios que participan en la conversación al calcular la heterogeneidad. La heterogeneidad de una red refleja la tendencia de contener *hub* y se define como $\sigma(k)/\text{mean}(k)$, en donde $\sigma(k)$ es la desviación estándar de los grados k y $\text{mean}(k)$ corresponde a la media de los grados k . Esta medida cuantifica la diversidad de usuarios según la proporción de seguidores versus amigos (Cardenas et al., 2018). También se observa un cambio en la heterogeneidad de los usuarios el día de la votación (ver el panel inferior de la figura 8.9 que muestra un aumento significativo. Esto significa que en el último día más usuarios diversos participaron en la conversación, y estos mismos usuarios causaron la disminución de la polarización.

8.5. Discusión

En este estudio se ha utilizado la metodología descrita en (Morales et al., 2015c) para analizar las elecciones presidenciales chilenas del 2017 y medir la polarización política resultante. Desde un pequeño número de usuarios influyentes, se ha estimado la opinión de la enorme mayoría, mostrando que durante el período precedente al día de la votación, la mayoría de los usuarios apoyaba a Piñera.

En la conversación de los usuarios de Twitter, se observa que estos usuarios desarrollan su opinión política y propagan los mensajes de su candidato preferido durante el período electoral, mostrando su decisión final el día anterior a la elección. En ese día, la polarización política alcanza su máximo valor cuando los usuarios ya decidieron su voto y limitan su comportamiento para apoyar a su líder y para amplificar los mensajes que lo respaldan. Este comportamiento propagandístico limita la discusión incrementando las opiniones radicales y la polarización. Después de este punto, la polarización política disminuyó durante el día de la votación, atrayendo a usuarios nuevos y más diversos que abrieron la conversación y aumentaron el flujo de información entre los dos lados.

Una alta correlación entre las opiniones de los usuarios el día 16 y el comportamiento de voto de la población chilena que enfrenta un enfoque dicotómico de las elecciones (candidato A o B) podría indicar que hasta ese día el proceso de formación de opinión de los usuarios de Twitter es racional y meditado, llegando a una decisión final de voto (vota A, B o no vota). Además esta decisión se consolida durante el día anterior a la votación,

donde los usuarios que participaron activamente durante la semana tienden a abandonar el debate y sólo apoyan su decisión, lo que provoca un pico máximo en la polarización política. Además, durante ese día, las opiniones inferidas de los usuarios de Twitter se correlacionan casi perfectamente con los resultados de las elecciones.

Ya que en este documento estimamos la opinión de un gran número de usuarios de Twitter, discutiremos las posibilidades y limitaciones de usar este método para predecir los resultados de las elecciones. Si asociamos la distribución de opiniones de los usuarios de Twitter durante la campaña electoral a sus intenciones de voto, podemos compararlas con el resultado final de las elecciones. Podemos asociar opiniones cercanas a $X = +1$ con la intención de votar a Piñera y valores de $X = -1$ con votantes potenciales de Guillier. En tal escenario, el día 16, el porcentaje de opiniones asociadas con Piñera es de alrededor del 54.4 % y las asociadas a Guillier alrededor del 45,6 %. El resultado fue una victoria de Piñera con el 54.5 % de los votos sobre Guillier con 45.4 % de los votos. Dado que los resultados coinciden con bastante precisión con los resultados finales, se podría pensar que esta metodología podría ser útil para inferir los resultados de las elecciones. Por lo tanto, a pesar de que los autores introdujeron en (Morales et al., 2015c) esta metodología para medir la polarización política, a la luz de estos resultados, la metodología también podría usarse para inferir los resultados de las elecciones. Sin embargo, para realizar esta tarea, se deberá hacer un análisis más profundo y tener en cuenta mas datos, como por ejemplo, ver qué perfiles están más o menos representados en Twitter. Por lo tanto, aún es necesario realizar nuevas investigaciones para extrapolar la metodología para predecir con eficacia los resultados electorales de las conversaciones en redes sociales.

Capítulo 9

Conclusiones

El objetivo de este trabajo fue modelizar diversos sistemas sociales a través de grafos, para poder utilizar las medidas de Actividad, Respuesta del Sistema y Centralidad de sus componentes en la construcción de diversos algoritmos e indicadores que permitan caracterizar este tipo de sistema, tanto presenciales como *online*.

A lo largo de esta tesis se han propuesto diversas

A continuación se definirán las conclusiones específicas de cada uno de los estudios.

1. Se ha mostrado que es posible utilizar los conceptos de Actividad, Respuesta de la Red y Centralidad en cualquier tipo de red que incluya interacciones humanas o incluso en aquellas que se derivan de estas interacciones, como son las redes de complemento disciplinar.
2. Cuando se trabaja con las redes de coautoría, se observa que para valores individuales se puede encontrar una fuerte correlación entre el Pagerank y el número de publicaciones, y a la vez una correlación algo menor, pero no menos importante, entre el Grado y el número de publicaciones. Esto da a entender que la posición estratégica dada por el Pagerank podrá favorecer al número de publicaciones de cada uno de los autores.
3. No se ha observado la presencia de correlación entre número de citas y el Pagerank, ni tampoco entre la eficiencia y el Pagerank.
4. Se homologó la metodología de construcción de redes de co-autoría para la construcción de redes de complemento disciplinar. Esto permitió entregar una serie de visualizaciones del conocimiento científico para cada uno de los catorce países presentados en el estudio, que sirvió de base para plantear una hipótesis respecto a que existe una diferencia en la generación del conocimiento entre los países desarrollados contra el resto de los países según la clasificación del Banco Mundial.

5. A través de redes de complemento disciplinar se desarrolló un indicador llamado Factor de Importancia Disciplinar (FID), el cual fue elaborado utilizando los conceptos de Actividad, Respuesta del Sistema y Centralidad. Éste índice entregó un listado de disciplinas, en donde aquellas que poseían un mayor FID, eran aquellas disciplinas científicas que colaboraban continuamente con otras y que además poseían un alto nivel de producción y de citas. Entre estas disciplinas se destacan *Multidisciplinary Sciences* y *Biochemistry & Molecular Biology*.
6. Utilizando el FID como base, se desarrolló un Índice de Complejidad Disciplinar (ICD), que es una medida que indica que tan alineado está el conocimiento de cada país, respecto a las disciplinas más importantes encontradas en el FID. Se muestra que el ICD se correlaciona fuertemente con la clasificación de países (desarrollados y en vías de desarrollo) otorgada por el Banco Mundial.
7. Por otro lado, el ICD se correlacionó con un coeficiente de Pearson de 0,82 con el PIB de cada uno de los países y, además se correlacionó con un coeficiente de Pearson de 0,84 con el dinero invertido en I+D con cada uno de los países del estudio.
8. A través de una Cartografía del Conocimiento, que consiste en una simplificación de las redes de complemento disciplinar, se desarrolló un algoritmo de predicción de disciplinas científicas que ayuden a guiar a las instituciones a desarrollar aquellas que no han sido desarrolladas. Este algoritmo entrega una alta Sensibilidad de los resultados, con un promedio de 0,84 para todas las universidades.
9. Para el caso de las redes de *Retweets*, se verificó que a través de una pequeña cantidad de usuarios de élite, cuya opinión era conocida, fue posible estimar la opinión política del resto de los usuarios oyentes.
10. Se observó que el punto máximo de polarización política se generó el día previo a las elecciones presidenciales chilenas del 2017. Por otro lado, la distribución de opinión de los usuarios en este día se correlacionó de manera casi perfecta con sus resultados electorales.
11. Finalmente, se observó un cambio en el comportamiento respecto a la opinión política de los usuarios de Twitter entre el período anterior a las elecciones presidenciales y el período que incluye las elecciones presidenciales, debido principalmente a la entrada de nuevos usuarios que participan en la discusión.

9.1. Trabajo a futuro

Como trabajo a futuro se plantea lo siguiente. Para el caso de la polarización en Twitter, es necesario realizar nuevas pruebas en otros países para comprobar si la correlación que se dió respecto a las elecciones presidenciales en Chile y los resultados electorales son producto de que la estructura de red que entrega Twitter es muy similar a la estructura de votantes de un país.

Por el lado de las redes de complemento disciplinar, es importante aumentar el número de países de estudio a modo de fortalecer el indicador. Además se plantea realizar pruebas adicionales y probar si la correlación se mantiene al comparar en ventanas temporales mas pequeñas.

Finalmente, es importante limpiar los datos de las citas para volver a compararlo con las medidas de Centralidad, esto debido a que, si se incluyen la edad de los artículos o de las personas, los resultados del estudio serán mucho mas precisos.

Bibliografía

- ADAMIC, L. A. y GLANCE, N. The political blogosphere and the 2004 u.s. election: Divided they blog. *Proceedings of the 3rd International Workshop on Link Discovery*, páginas 36–43, 2005.
- AKSOY, S., KOLDA, T. y PINAR, A. Measuring and modeling bipartite graphs with community structure. *Journal of Complex Networks*, vol. 5(4), páginas 581–603, 2016.
- AMERICAECONOMÍA. Ránking de universidades de chile 2018. <https://www.americaeconomia.com/negocios-industrias/estas-son-las-500-empresas-mas-grandes-de-chile-2019>. Última visita: 26/08/2019, 2018.
- ASSENOV, Y., RAMÍREZ, F., SCHELHORN, S., LENGAUER, T. y ALBRECHT, M. Computing topological parameters of biological networks. *Bioinformatics*, vol. 24(2), páginas 282–4, 2008.
- BAGLER, G. y SINHA, S. Assortative mixing in protein contact networks and protein folding kinetics. *Bioinformatics*, vol. 23(14), páginas 1760–1767, 2007.
- BAR-YAM, Y. General features of complex systems. *Encyclopedia of Life Support Systems*, vol. I, 2002.
- BECKER, M. Y. y ROJAS, I. A graph layout algorithm for drawing metabolic pathways. *Bioinformatics*, vol. 17(5), páginas 461–467, 2001.
- BLONDEL, V., GAJARDO, A., HEYMANS, M., SENELLART, P. y DOOREN, P. V. A measure of similarity between graph vertices: Applications to synonym extraction and web searching. *SIAM Review*, vol. 46(4), páginas 647–666, 2004.
- BORONDO, J. *Aplicación de la teoría de redes complejas a procesos dinámicos en la sociedad*. Tesis Doctoral, Universidad Politécnica de Madrid, 2015.
- BORONDO, J., BORONDO, F., RODRIGUEZ-SICKERT, C. y HIDALGO, C. A. To each according to its degree: The meritocracy and topocracy of embedded markets. *Scientific Reports*, vol. 3784, 2014.

- BORONDO, J., MORALES, A. J., BENITO, R. M. y LOSADA, J. C. Multiple leaders on a multilayer social media. *Chaos, Solitons & Fractals*, vol. 72, páginas 90–98, 2015.
- BORONDO, J., MORALES, A. J., LOSADA, J. C. y BENITO, R. M. Characterizing and modeling an electoral campaign in the context of twitter: 2011 spanish presidential election as a case study. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 22, 2012.
- BOYD, D., GOLDER, S. y LOTAN, G. Tweet, tweet, retweet: conversational aspects of retweeting on twitter. *Proceedings of the 43rd Annual Hawaii International Conference on System Sciences (HICSS '10)*, páginas 51–60, 2010.
- BRANDES, U. A faster algorithm for betweenness centrality. *J Math Sociol*, vol. 25(163-177), páginas 3106–3108, 2001.
- CALLAWAY, D., HOPCROFT, J., KLEINBERG, J., NEWMAN, M. E. y S.H.STROGATZ. Are randomly grown graphs really random? *Physical Review E*, vol. 64(4), 2001.
- CARDENAS, J., CABRERA, F., MOGUILLANSKY, G. y OLIVARES, G. Cartografía del conocimiento en chile. *CNID, Chile*, 2015.
- CARDENAS, J. P., VIDAL, G., URBINA, C., OLIVARES, G., RODRIGO, P. y FUENTES, M. Social crises: signatures of complexity in a fast- growing economy. *Complexity*, 2018.
- CLEMENS, M. Visualizing complex systems science (css). <https://necsi.edu/visualizing-complex-systems-science>. Última visita: 26/08/2019, 2019.
- CONOVER, M. D., GONÇALVES, B., FLAMMINI, A. y MENCZER, F. Partisan asymmetries in online political activity. *EPJ Data Science*, vol. 1(6), páginas 1–17, 2012.
- CONOVER, M. D., RATKIEWICZ, J., FRANCISCO, M., GONC, B., FLAMMINI, A. y MENCZER, F. Political polarization on twitter. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, páginas 89–96, 2011.
- CREIXELL, W., LOSADA, J., ARREDONDO, T., OLIVARES, P. y BENITO., R. Serendipity in social network. *NETWORKS AND HETEROGENEOUS MEDIA*, vol. 7(3), páginas 363–371, 2012.
- CRUZ-COKE, R. Evolucion de las universidades chilenas 1981-. *Revista medica de Chile*, vol. 132(12), páginas 1543–1549, 2004.

- DEY, A. Machine learning algorithms: A review. *International Journal of Computer Science and Information Technologies*, vol. 7(3), páginas 1174–1179, 2016.
- DIESTEL, R. Graph theory. *Springer-Verlag*, 2005.
- ERGÜN, G. Human sexual contact network as a bipartite graph. *arXiv*, 2001.
- GONZALEZ-BRAMBILA, C., REYES-GONZALEZ, L., VELOSO, F. y ANGON, M. P. The scientific impact of developing nations. *Plos One*, vol. 11(3), 2016.
- GONÇALVES, B., PERRA, N., VESPIGNANI, A. y PERC, M. Modeling user-sactivity on twitter networks: validation of dunbar’s number. *Plos One*, vol. 6(8), 2011.
- GRAHAM, R. y HELL, P. On the history of the minimum spanning tree problem. 1985.
- GUIMERA, R., MOSSA, S., TURTSCHI, A. y AMARAL, L. A. N. The world-wide air transportation network: Anomalous centrality, community structure, and cities global roles. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102(22), páginas 7794–7799, 2005.
- HARRINGTON, P. Visualización de los datos antes y despues de aplicar pca. *Machine Learning in action*, 2012.
- HERDAGDELEN, A., ZUO, W., GARD-MURRAY, A. y BAR-YAM, Y. An exploration of social identity: The geography and politics of news-sharing communities in twitter. *Complexity*, vol. 19(2), páginas 10–20, 2013.
- HERRERA-YAGÜE, C., SCHNEIDER, C. M., COURONNÉ, T., SMOREDA, Z., BENITO, R. M., ZUFIRIA, P. J. y GONZÁLEZ, M. C. The anatomy of urban social networks and its implications in the searchability problem. *Scientific Reports* 5, 2015.
- HIDALGO, C., KLINGER, B., BARABASI, A. y HAUSMANN, R. The product space conditions the development of nations. *Science*, vol. 317, páginas 482–487, 2007.
- HIDALGO, C. A. y HAUSMANN, R. The building blocks of economic complexity. *PNAS*, vol. 106(26), páginas 10570–10575, 2009.
- HIPERTEXTUAL. ¿qué es una api? <https://hipertextual.com/archivo/2014/05/que-es-api/>. Última visita: 26/08/2019, 2014.
- HUBERMAN, B. A., ROMERO, D. M. y WU, F. Social networks that matter twitter under the microscope. *First Monday*, vol. 14(1), 2009.

- INTERNACIONAL, F. E. World economic outlook database. 2019.
- JUNGHERR, A. Tweets and votes, a special relationship: The 2009 federal election in germany. *Proceedings of the 2nd Workshop on Politics, Elections and Data, PLEAD 2013*, páginas 5–14, 2013.
- KANUNGO, T., NETANYAHU, N. S. y WU, A. Y. An efficient k-means clustering algorithm: Analysis and implementation. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 24(7), 2002.
- KENDALL, M. G. y STUART, A. *The Advanced Theory of Statistics, Volume 2: Inference and Relationship..* Griffin, 1973.
- KOLJATIC, M. y SILVA, M. The international publication productivity of latin american countries in the economics and business administration fields. *Scientometrics*, vol. 51(2), páginas 381–394, 2001.
- KOTSIANTIS, S. B. Supervised machine learning: A review of classification techniques. *Informatica*, vol. 31, páginas 249–268, 2007.
- LAUTERBACH, R. Chile y la complejidad de sus exportaciones. *CNID, Chile*, 2016.
- MARIN, J. Análisis de componentes principales. 2008.
- MARTIN-GUTIERREZ, S., BORONDO, J., MORALES, A., LOSADA, J., TARQUIS, A. y BENITO., R. Agricultural activity shapes the communication and migration patterns in senegal. *CHAOS*, vol. 26(6), 2016.
- MARTIN-GUTIERREZ, S., LOSADA, J. y BENITO, R. Recurrent patterns of user behavior in different electoral campaigns: A twitter analysis of the spanish general elections of 2015 and 2016. *Complexity*, 2018.
- MEYER, D. y WIEN, T. U. Support vector machines. the interface to libsvm in package e1071. online-documentation of the package e1071 for r. 2001.
- MILGRAM, S. The small world problem. *Psychology Today*. Ziff-Davis Publishing Company., 1967.
- MORALES, A., BORONDO, J., LOSADA, J. y BENITO, R. Efficiency of human activity on information spreading on twitter. *Social Networks*, 2015a.
- MORALES, A., CREIXELL, W., BORONDO, J., LOSADA, J. y BENITO., R. Characterizing ethnic interactions from human communication patterns in ivory coast. *NETWORKS AND HETEROGENEOUS MEDIA*, vol. 10(1), páginas 87–99, 2015b.

- MORALES, A. J., BORONDO, J., LOSADA, J. C. y BENITO, R. M. Measuring political polarization: Twitter shows the two sides of venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, 2015c.
- MORALES, A. J., LOSADA, J. C. y BENITO, R. M. Users structure and behavior on an online social network during a political protest. *Physica A: Statistical Mechanics and its Applications*, vol. 391(21), páginas 5244–5253, 2012.
- MURRAY, G.-M. El quark y el jaguar. *Tusquets Editores*, 1995.
- NEWMAN, M. A measure of betweenness centrality based on random walks. *arXiv*, 2003.
- NEWMAN, M. *Networks. An Introduction*. Oxford, 2010.
- OLIVARES, G., CÁRDENAS, J., LOSADA, J. y BORONDO, J. Opinion polarization during a dichotomous electoral process. *Complexity*, 2019.
- ORCID. Orcid. <https://orcid.org/>. Última visita: 26/08/2019, 2018.
- PAGE, L., BRIN, S., MOTWANI, R. y WINOGRAD, T. The pagerank citation ranking: Bringing order to the web. *Technical Report*, 1998.
- PARMELEE, J. H. Political journalists and twitter: Influences on norms and practices. *Journal of Media Practice*, vol. 14(4), páginas 291–305, 2013.
- PATRO, S. G. K. y SAHU, K. K. Normalization: A preprocessing stage. *arXiv*, 2015. DOI: 10.17148/IARJSET.2015.2305.
- RCN-RADIO. Twitter contabiliza más de 200 millones de tuits sobre «brexit» desde 2016. <https://www.rcnradio.com/internacional/twitter-contabiliza-mas-de-200-millones-de-tuits-sobre-brexit-desde-2016>. Última visita: 26/08/2019, 2019.
- SAMUEL, A. Algunos estudios en aprendizaje automático utilizando el juego de damas. *Revista IBM de Investigación y Desarrollo*, vol. 3, páginas 210–229, 1959.
- SERRA, G. Should we expect primary elections to create polarization? a robust median voter theorem with rational parties. *Routledge Handbook of Primary Elections*, R. G. Boatright, 2018.
- SIMON, H. A. How complex are complex systems. *Philosophy of Science Associatio*, 1977.
- SOLOP, F. I. Rt @barack obama we just made history: Twitter and the 2008 presidential election. *Communicator-In-Chief: How Barack Obama Used New Media Technology to Win The White House*, J. A. Hendricks, J. Robert, and E. Denton, *Lexicon Books*, páginas 37–49, 2010.

- SRIVASTAVA, D. K. y BHAMBHU, L. Data classification using support vector machine. *Journal of Theoretical and Applied Information Technology*, vol. 12(1), páginas 1–7, 2010.
- STAFFORD-FRASER, Q. The life and times of the first web cam. <https://www.cl.cam.ac.uk/coffee/qsf/cacm200107.html>. Última visita: 26/08/2019, 2001.
- SUNDQUIST, J. L. Dynamics of the party system: Alignment and realignment of political parties in the united states. *The Brookings Institution, Washington*, 1983.
- SUÁREZ, E. J. C. Tutorial sobre máquinas de vectores soporte (svm). 2014.
- TORRES, J. Introducción a los sistemas complejos. 2010.
- TWITTER. Cómo usar los hashtags. <https://help.twitter.com/es/using-twitter/how-to-use-hashtags>. Última visita: 26/08/2019, 2011.
- WALDROP, M. *The emerging science at the edge of the order and chaos*. Simon & Schuster, 1992.
- WANG, Y., FENG, Y., HONG, Z., BERGER, R. y LUO, J. How Polarized Have We Become? A Multimodal Classification of Trump Followers and Clinton Followers. *arXiv e-prints*, página arXiv:1711.00617, 2017.
- WASSERMAN, S. y FAUST, K. Social network analysis: Methods and applications. *Cambridge University Press*, 1994.
- WATTS, D. J. y STROGATZ, S. H. Collective dynamics of small-world networks. *Nature*, vol. 393, páginas 440–442, 1998.
- YAQUB, U., CHUN, S. A., ATLURI, V. y VAIDYA, J. Analysis of political discourse on twitter in the context of the 2016 us presidential elections. *Government Information Quarterly*, vol. 34(4), páginas 613–626, 2017.
- YOON, J., BLUMER, A. y LEE, K. An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality. *Bioinformatics*, vol. 22, páginas 3106–3108, 2006.
- YU, L. y LIU, H. Efficient feature selection via analysis of relevance and redundancy. *JMLR*, vol. 5, páginas 1205–1224, 2004.
- ZACHARY, W. W. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research*, vol. 33, 1977.
- ZAIDI, F. Small world networks and clustered small world networks with random connectivity. *Social Network Analysis and Mining*, 2012.

Parte IV

Apéndices

Apéndice A

Aprendizaje automático

A.1. Introducción al aprendizaje automático

Los seres humanos suelen cometer errores al realizar análisis, especialmente al establecer relaciones entre diversos datos, lo que dificulta encontrar soluciones a muchos problemas, por esta razón el uso de algoritmos de aprendizaje automático sirven como herramientas de apoyo para encontrar patrones entre estos datos.

El término de aprendizaje automático fue acuñado en 1959 por Samuel Arthur (Samuel, 1959) y está basado en una serie de algoritmos y modelos estadísticos que permiten a los sistemas informáticos realizar tareas específicas a través de un conjunto de datos (características) para dar respuesta a una problemática a resolver.

Un conjunto de datos se puede ejemplificar en la tabla A.1. En esta tabla se observa un conjunto de datos formado por N elementos y con Z características en donde cada uno tiene una clase asociada que puede ser verdadero o falso. Esta es una representación abstracta de lo que podría ser la clasificación de, por ejemplo, resultado final de un grupo de estudiantes de un curso dado, en donde la clase sería Aprobado o No Aprobado y sus características podrían ser sus notas, ingresos, asistencia, entre otras. A continuación se describirán dos tipos de algoritmos de aprendizaje, que son los que se han utilizado a lo largo de esta tesis y son los algoritmos de Aprendizaje

Número	Característica A	Característica B	...	Característica Z	Clase
1	A1	B1	...	Z1	Verdadero
2	A2	B2	...	Z2	Falso
...	...	...	...	...	...
N	AN	BN	...	ZN	Verdadero

Tabla A.1: Ejemplo de conjunto de datos para ser utilizados en aprendizaje automático

Supervisado y los de Aprendizaje no Supervisado.

A.2. Aprendizaje Supervisado

El aprendizaje automático supervisado es aquel proceso en que, a través de un conjunto de reglas dado, es posible realizar inferencias automatizadas a través de los datos, con ayuda de la supervisión humana (Kotsiantis, 2007). Estos algoritmos suelen ser utilizados especialmente para clasificación y regresión. En la figura A.1 se puede ver el esquema básico del aprendizaje automático supervisado, en donde el primer paso consiste en la recolección de un conjunto de datos para resolver un problema específico. Es importante que exista la supervisión de un experto que pueda ayudar a sugerir cuales son aquellos atributos o características más significativos, puesto que sin la ayuda humana, el sistema entraría pasaría simplemente a resolver un problema de fuerza bruta.

Figura A.1: Diagrama de flujo de un algoritmo de aprendizaje automático supervisado.

En el siguiente paso se procede al pre-procesamiento de los datos, en donde, dependiendo de la naturaleza del problema, se deben aplicar diversas técnicas de limpieza y/o normalización de datos, de manera tal que se pueda eliminar la mayor cantidad de ruido posible.

Para este caso, la normalización de ciertos parámetros puede ser muy importante para comparar valores entre diversos actores, ya que de esta forma podemos comparar a igualdad de escala. Por otro lado, la limpieza de los datos puede ayudar significativamente para el desarrollo de un mejor modelo, ya que en muchas ocasiones los datos pueden estar mal etiquetados, errores simples como tener tabulada una celda con el valor «Verdadero», mientras que en otra por error se escribió «Verdaderos» (una s adicional al final del texto), lo que provoca que el sistema los identifique como distintos valores.

El tercer paso consiste en la definición de los datos de entrenamiento y eliminar todos aquellos conjuntos de datos que sean irrelevantes para el sistema (Yu y Liu, 2004). Esto permite reducir la dimensionalidad de los datos y permite que los algoritmos de aprendizaje automático supervisado funcionen más rápidos y eficientes.

El siguiente paso consiste en la selección del algoritmo de aprendizaje automático, en donde, para el caso de un clasificador, la calidad del algoritmo dependerá de la precisión de la predicción, es decir, el porcentaje de aciertos partido por el número total de predicciones.

Generalmente para la selección del conjunto de datos de entrenamiento se suele utilizar dos tercios de los datos y un tercio para la evaluación para estimar el rendimiento.

Tanto en el proceso de entrenamiento y como en el de evaluación se puede ir retroalimentando esta información al sistema para obtener mejores resultados, para finalmente, dar por terminado el procedimiento en el caso de obtener los valores deseados.

Dentro de los algoritmos de aprendizaje automático supervisado destacaremos las máquinas de soporte vectorial, las que se tratarán a continuación.

A.2.1. Máquinas de Soporte Vectorial

Las máquinas de soporte vectorial (SVM) son uno de los algoritmos de aprendizaje automático supervisado más utilizados, las que se utilizan principalmente para la clasificación binaria o regresión (Dey, 2016). Su principio radica principalmente en dibujar líneas en el espacio de datos de manera tal que la distancia entre el margen y las clases que hemos definido previamente se minimice a través de los llamados vectores de soporte, tal como se aprecia en la figura A.2.

Mientras la mayoría de los métodos de aprendizaje se centran en minimizar los errores cometidos por el modelo generado a partir de los ejemplos de en-

Figura A.2: Funcionamiento de una máquina de soporte vectorial (Meyer y Wien, 2001).

trenamiento (error empírico), el sesgo inductivo asociado a las SVMs radica en la minimización del denominado riesgo estructural. La idea es seleccionar un hiperplano de separación que equidista de los puntos más cercanos de cada clase para, de esta forma, conseguir lo que se denomina un margen máximo a cada lado del hiperplano. Además, a la hora de definir el hiperplano, sólo se consideran los puntos de entrenamiento de cada clase que caen justo en la frontera de dichos márgenes a los que se les llama vectores de soporte. Desde un punto de vista práctico, el hiperplano separador de margen máximo ha demostrado tener una buena capacidad de generalización, evitando en gran medida el problema del sobreajuste a los ejemplos de entrenamiento (Suárez, 2014).

A.3. Aprendizaje No Supervisado

Los algoritmos de aprendizaje no supervisado aprenden algunas características de un set de datos, de esta forma es capaz de reconocer a qué clases corresponden los datos. Este tipo de algoritmos suelen ser útiles para análisis de agrupamiento de datos, que consiste en la clasificación de *clusters* o subgrupos según ciertos criterios.

Dentro de los algoritmos de aprendizaje automático no supervisados destacaremos el análisis de componentes principales y el k-medios, los que se tratarán a continuación.

A.3.1. Análisis de Componentes Principales

El análisis de componentes principales (PCA) nace de la idea en que cuando se recoge información de una muestra de datos, es frecuente tomar el mayor número de variables posibles, pero el problema es que al tener un número m de variables, habrá que considerar la combinatoria de esos m elementos para calcular sus posibles coeficientes de correlación. Por esta razón reducir el número de variables se vuelve una tarea necesaria.

Figura A.3: Visualización de los datos antes y después de aplicar PCA (Harrington, 2012)

La idea principal consiste en considerar un número m de variables correlacionadas para posteriormente transformar el conjunto original de datos a otro conjunto de nuevas variables, como se muestra en la figura A.3, a lo cual se le denomina componentes principales.

Estas nuevas variables son combinaciones lineales de las anteriores y se construyen según el orden de importancia en cuanto a la variabilidad total que recogen la muestra.

Para el cálculo de los componentes principales se considerará una serie de variables $(x_1, x_2, \dots, x_p)$ sobre un grupo de objetos para tratar de calcular un nuevo conjunto de variables $(y_1, y_2, \dots, y_3)$, que son combinación lineal de los datos iniciales (Marin, 2008) como se muestra en la ecuación A.1.

$$y_j = a_{j1}x_1 + a_{j2}x_2 + \dots + a_{jp}x_p = \mathbf{a}'_j \mathbf{x} \quad (\text{A.1})$$

donde p es el número de variables y $\mathbf{a}'_j$ un vector de constantes y $\mathbf{x}$ el vector de datos originales. Si se quisiera maximizar la varianza de nuestro nuevo vector $\mathbf{a}'$ lo que se haría es calcular la ecuación A.2.

$$\mathbf{a}'_j \mathbf{a}_j = \sum_{k=1}^p a_{kj}^2 = 1 \quad (\text{A.2})$$

A.3.2. K-Medios

El algoritmo de K-medios es un método de agrupamiento no supervisado. Este algoritmo necesita como dato de entrada el número de grupos en que se segmentarán los datos, de esa forma obtenemos *km clusters* con *km* puntos asignados de forma aleatoria, también llamado centroide (Kanungo et al., 2002). Se puede observar el diagrama de flujo del algoritmo en la figura A.4. El algoritmo continúa moviendo los centroides y calcula la distancia

Figura A.4: Diagrama de flujo del algoritmo k-medios.

mínima, utilizando, por ejemplo la distancia euclidiana. Así, se asigna cada objeto a cada grupo teniendo en cuenta el mínimo error de partición de cada elemento. De esta forma, al mover los centroides los grupos se mantienen estables, es decir, los elementos no son reasignados de grupo, entonces el algoritmo finaliza y obtenemos un conjunto de grupos.

Apéndice B

Visualización de redes de complemento disciplinar

En este apéndice se presentarán las visualizaciones de las redes de complemento disciplinar correspondientes a los tres niveles disciplinares, entre los que encontramos el nivel disciplinar, nivel de sub-categoría y finalmente el nivel de áreas de la OECD para cada uno de los países presentados en el capítulo 7.

B.1. Disciplinas correspondientes a cada nodo

En esta subsección se mostrarán las disciplinas que corresponden a cada uno de los nodos de cada una de las visualizaciones, los que se mostrarán a través de figuras con números a un costado de cada nodo y una tabla de conversión correspondiente.

Figura B.1: Identificador para cada una de las disciplinas dentro de la visualización.

Área OECD		Disciplina	Identificador
AGRICULTURAL SCIENCES		Veterinary Sciences	1
AGRICULTURAL SCIENCES		Agronomy	2
AGRICULTURAL SCIENCES		Agriculture, Multidisciplinary	3
AGRICULTURAL SCIENCES		Soil Science	4
AGRICULTURAL SCIENCES		Agriculture, Dairy & Animal Science	5
AGRICULTURAL SCIENCES		Fisheries	6
AGRICULTURAL SCIENCES		Forestry	7
AGRICULTURAL SCIENCES		Agricultural Engineering	8
AGRICULTURAL SCIENCES		Horticulture	9
AGRICULTURAL SCIENCES		Agricultural Economics & Policy	10
ENGINEERING TECHNOLOGY	AND	Materials Science, Paper & Wood	1
ENGINEERING TECHNOLOGY	AND	Engineering, Aerospace	2
ENGINEERING TECHNOLOGY	AND	Engineering, Marine	3
ENGINEERING TECHNOLOGY	AND	Nanoscience & Nanotechnology	4
ENGINEERING TECHNOLOGY	AND	Materials Science, Multidisciplinary	5
ENGINEERING TECHNOLOGY	AND	Engineering, Multidisciplinary	6
ENGINEERING TECHNOLOGY	AND	Engineering, Chemical	7
ENGINEERING TECHNOLOGY	AND	Energy & Fuels	8
ENGINEERING TECHNOLOGY	AND	Materials Science, Coatings & Films	9
ENGINEERING TECHNOLOGY	AND	Engineering, Biomedical	10
ENGINEERING TECHNOLOGY	AND	Remote Sensing	11

Área OECD		Disciplina	Identificador
ENGINEERING TECHNOLOGY	AND	Imaging Science & Photographic Technology	12
ENGINEERING TECHNOLOGY	AND	Mechanics	13
ENGINEERING TECHNOLOGY	AND	Biotechnology & Applied Microbiology	14
ENGINEERING TECHNOLOGY	AND	Food Science & Technology	15
ENGINEERING TECHNOLOGY	AND	Computer Science, Hardware & Architecture	16
ENGINEERING TECHNOLOGY	AND	Engineering, Electrical & Electronic	17
ENGINEERING TECHNOLOGY	AND	Metallurgy & Metallurgical Engineering	18
ENGINEERING TECHNOLOGY	AND	Telecommunications	19
ENGINEERING TECHNOLOGY	AND	Engineering, Industrial	20
ENGINEERING TECHNOLOGY	AND	Automation & Control Systems	21
ENGINEERING TECHNOLOGY	AND	Materials Science, Biomaterials	22
ENGINEERING TECHNOLOGY	AND	Robotics	23
ENGINEERING TECHNOLOGY	AND	Construction & Building Technology	24
ENGINEERING TECHNOLOGY	AND	Microscopy	25
ENGINEERING TECHNOLOGY	AND	Engineering, Petroleum	26
ENGINEERING TECHNOLOGY	AND	Instruments & Instrumentation	27
ENGINEERING TECHNOLOGY	AND	Nuclear Science & Technology	28
ENGINEERING TECHNOLOGY	AND	Thermodynamics	29
ENGINEERING TECHNOLOGY	AND	Engineering, Mechanical	30
ENGINEERING TECHNOLOGY	AND	Spectroscopy	31
ENGINEERING TECHNOLOGY	AND	Engineering, Civil	32

Área OECD		Disciplina	Identificador
ENGINEERING TECHNOLOGY	AND	Engineering, Geological	33
ENGINEERING TECHNOLOGY	AND	Engineering, Environmental	34
ENGINEERING TECHNOLOGY	AND	Cell & Tissue Engineering	35
ENGINEERING TECHNOLOGY	AND	Medical Laboratory Technology	36
ENGINEERING TECHNOLOGY	AND	Materials Science, Characterization & Testing	37
ENGINEERING TECHNOLOGY	AND	Engineering, Manufacturing	38
ENGINEERING TECHNOLOGY	AND	Materials Science, Composites	39
ENGINEERING TECHNOLOGY	AND	Materials Science, Ceramics	40
ENGINEERING TECHNOLOGY	AND	Transportation Science & Technology	41
ENGINEERING TECHNOLOGY	AND	Mining & Mineral Processing	42
ENGINEERING TECHNOLOGY	AND	Engineering, Ocean	43
ENGINEERING TECHNOLOGY	AND	Materials Science, Textiles	44
HUMANITIES		Literature, American	1
HUMANITIES		Poetry	2
HUMANITIES		Folklore	3
HUMANITIES		Literature, British Isles	4
HUMANITIES		Medieval & Renaissance Studies	5
HUMANITIES		Classics	6
HUMANITIES		Art	7
HUMANITIES		Literature	8
HUMANITIES		Literary Theory & Criticism	9
HUMANITIES		Literature, African, Australian, Canadian	10
HUMANITIES		Literature, German, Dutch, Scandinavian	11
HUMANITIES		Theater	12
HUMANITIES		Literary Reviews	13
HUMANITIES		Literature, Slavic	14
HUMANITIES		Archaeology	15
HUMANITIES		Architecture	16

Área OECD	Disciplina	Identificador
HUMANITIES	Humanities, Multidisciplinary	17
HUMANITIES	History	18
HUMANITIES	Philosophy	19
HUMANITIES	History & Philosophy Of Science	20
HUMANITIES	Linguistics	21
HUMANITIES	Language & Linguistics	22
HUMANITIES	Religion	23
HUMANITIES	Ethics	24
HUMANITIES	Film, Radio, Television	25
HUMANITIES	Literature, Romance	26
HUMANITIES	Music	27
HUMANITIES	History Of Social Sciences	28
MEDICAL AND HEALTH SCIENCES	Psychology, Psychoanalysis	1
MEDICAL AND HEALTH SCIENCES	Medicine, General & Internal	2
MEDICAL AND HEALTH SCIENCES	Psychiatry	3
MEDICAL AND HEALTH SCIENCES	Peripheral Vascular Disease	4
MEDICAL AND HEALTH SCIENCES	Chemistry, Medicinal	5
MEDICAL AND HEALTH SCIENCES	Nursing	6
MEDICAL AND HEALTH SCIENCES	Transplantation	7
MEDICAL AND HEALTH SCIENCES	Ophthalmology	8
MEDICAL AND HEALTH SCIENCES	Public, Environmental & Occupational Health	9
MEDICAL AND HEALTH SCIENCES	Medicine, Research & Experimental	10
MEDICAL AND HEALTH SCIENCES	Nutrition & Dietetics	11
MEDICAL AND HEALTH SCIENCES	Clinical Neurology	12
MEDICAL AND HEALTH SCIENCES	Parasitology	13
MEDICAL AND HEALTH SCIENCES	Tropical Medicine	14
MEDICAL AND HEALTH SCIENCES	Dentistry, Oral Surgery & Medicine	15

Área OECD	Disciplina	Identificador
MEDICAL AND HEALTH SCIENCES	Respiratory System	16
MEDICAL AND HEALTH SCIENCES	Pediatrics	17
MEDICAL AND HEALTH SCIENCES	Infectious Diseases	18
MEDICAL AND HEALTH SCIENCES	Critical Care Medicine	19
MEDICAL AND HEALTH SCIENCES	Pharmacology & Pharmacy	20
MEDICAL AND HEALTH SCIENCES	Immunology	21
MEDICAL AND HEALTH SCIENCES	Allergy	22
MEDICAL AND HEALTH SCIENCES	Endocrinology & Metabolism	23
MEDICAL AND HEALTH SCIENCES	Neurosciences	24
MEDICAL AND HEALTH SCIENCES	Obstetrics & Gynecology	25
MEDICAL AND HEALTH SCIENCES	Urology & Nephrology	26
MEDICAL AND HEALTH SCIENCES	Radiology, Nuclear Medicine & Medical Imaging	27
MEDICAL AND HEALTH SCIENCES	Toxicology	28
MEDICAL AND HEALTH SCIENCES	Rheumatology	29
MEDICAL AND HEALTH SCIENCES	Pathology	30
MEDICAL AND HEALTH SCIENCES	Anesthesiology	31
MEDICAL AND HEALTH SCIENCES	Cardiac & Cardiovascular Systems	32
MEDICAL AND HEALTH SCIENCES	Medicine, Legal	33
MEDICAL AND HEALTH SCIENCES	Hematology	34
MEDICAL AND HEALTH SCIENCES	Oncology	35
MEDICAL AND HEALTH SCIENCES	Integrative & Complementary Medicine	36

Área OECD	Disciplina	Identificador
MEDICAL AND HEALTH SCIENCES	Surgery	37
MEDICAL AND HEALTH SCIENCES	Gastroenterology & Hepatology	38
MEDICAL AND HEALTH SCIENCES	Orthopedics	39
MEDICAL AND HEALTH SCIENCES	Dermatology	40
MEDICAL AND HEALTH SCIENCES	Medical Ethics	41
MEDICAL AND HEALTH SCIENCES	Social Sciences, Biomedical	42
MEDICAL AND HEALTH SCIENCES	Physiology	43
MEDICAL AND HEALTH SCIENCES	Psychology, Clinical	44
MEDICAL AND HEALTH SCIENCES	Neuroimaging	45
MEDICAL AND HEALTH SCIENCES	Emergency Medicine	46
MEDICAL AND HEALTH SCIENCES	Health Care Sciences & Services	47
MEDICAL AND HEALTH SCIENCES	Health Policy & Services	48
MEDICAL AND HEALTH SCIENCES	Anatomy & Morphology	49
MEDICAL AND HEALTH SCIENCES	Primary Health Care	50
MEDICAL AND HEALTH SCIENCES	Geriatrics & Gerontology	51
MEDICAL AND HEALTH SCIENCES	Gerontology	52
MEDICAL AND HEALTH SCIENCES	Medical Informatics	53
MEDICAL AND HEALTH SCIENCES	Otorhinolaryngology	54
MEDICAL AND HEALTH SCIENCES	Andrology	55
MEDICAL AND HEALTH SCIENCES	Sport Sciences	56
MEDICAL AND HEALTH SCIENCES	Substance Abuse	57

Área OECD	Disciplina	Identificador
MEDICAL AND HEALTH SCIENCES	Audiology & Speech-Language Pathology	58
MEDICAL AND HEALTH SCIENCES	Rehabilitation	59
NATURAL SCIENCES	Ornithology	1
NATURAL SCIENCES	Physics, Condensed Matter	2
NATURAL SCIENCES	Chemistry, Multidisciplinary	3
NATURAL SCIENCES	Chemistry, Physical	4
NATURAL SCIENCES	Physics, Applied	5
NATURAL SCIENCES	Entomology	6
NATURAL SCIENCES	Zoology	7
NATURAL SCIENCES	Geosciences, Multidisciplinary	8
NATURAL SCIENCES	Astronomy & Astrophysics	9
NATURAL SCIENCES	Physics, Particles & Fields	10
NATURAL SCIENCES	Genetics & Heredity	11
NATURAL SCIENCES	Physics, Multidisciplinary	12
NATURAL SCIENCES	Biochemistry & Molecular Biology	13
NATURAL SCIENCES	Multidisciplinary Sciences	14
NATURAL SCIENCES	Optics	15
NATURAL SCIENCES	Physics, Atomic, Molecular & Chemical	16
NATURAL SCIENCES	Environmental Sciences	17
NATURAL SCIENCES	Meteorology & Atmospheric Sciences	18
NATURAL SCIENCES	Plant Sciences	19
NATURAL SCIENCES	Chemistry, Organic	20
NATURAL SCIENCES	Ecology	21
NATURAL SCIENCES	Computer Science, Software Engineering	22
NATURAL SCIENCES	Geography, Physical	23
NATURAL SCIENCES	Physics, Nuclear	24
NATURAL SCIENCES	Physics, Mathematical	25
NATURAL SCIENCES	Geology	26
NATURAL SCIENCES	Statistics & Probability	27
NATURAL SCIENCES	Computer Science, Interdisciplinary Applications	28
NATURAL SCIENCES	Microbiology	29
NATURAL SCIENCES	Biophysics	30
NATURAL SCIENCES	Limnology	31
NATURAL SCIENCES	Marine & Freshwater Biology	32
NATURAL SCIENCES	Oceanography	33

Área OECD	Disciplina	Identificador
NATURAL SCIENCES	Computer Science, Theory & Methods	34
NATURAL SCIENCES	Chemistry, Applied	35
NATURAL SCIENCES	Cell Biology	36
NATURAL SCIENCES	Mathematics, Interdisciplinary Applications	37
NATURAL SCIENCES	Biology	38
NATURAL SCIENCES	Virology	39
NATURAL SCIENCES	Computer Science, Artificial Intelligence	40
NATURAL SCIENCES	Mathematics	41
NATURAL SCIENCES	Chemistry, Analytical	42
NATURAL SCIENCES	Mathematics, Applied	43
NATURAL SCIENCES	Polymer Science	44
NATURAL SCIENCES	Biodiversity Conservation	45
NATURAL SCIENCES	Evolutionary Biology	46
NATURAL SCIENCES	Mycology	47
NATURAL SCIENCES	Chemistry, Inorganic & Nuclear	48
NATURAL SCIENCES	Acoustics	49
NATURAL SCIENCES	Paleontology	50
NATURAL SCIENCES	Geochemistry & Geophysics	51
NATURAL SCIENCES	Crystallography	52
NATURAL SCIENCES	Electrochemistry	53
NATURAL SCIENCES	Logic	54
NATURAL SCIENCES	Water Resources	55
NATURAL SCIENCES	Mathematical & Computational Biology	56
NATURAL SCIENCES	Biochemical Research Methods	57
NATURAL SCIENCES	Computer Science, Information Systems	58
NATURAL SCIENCES	Reproductive Biology	59
NATURAL SCIENCES	GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY	60
NATURAL SCIENCES	Physics, Fluids & Plasmas	61
NATURAL SCIENCES	Developmental Biology	62
NATURAL SCIENCES	Mineralogy	63
NATURAL SCIENCES	Computer Science, Cybernetics	64
SOCIAL SCIENCES	Cultural Studies	1
SOCIAL SCIENCES	Psychology, Experimental	2
SOCIAL SCIENCES	Ergonomics	3

Área OECD	Disciplina	Identificador
SOCIAL SCIENCES	Ethnic Studies	4
SOCIAL SCIENCES	Women's Studies	5
SOCIAL SCIENCES	Psychology, Social	6
SOCIAL SCIENCES	Family Studies	7
SOCIAL SCIENCES	Asian Studies	8
SOCIAL SCIENCES	Education & Educational Research	9
SOCIAL SCIENCES	Urban Studies	10
SOCIAL SCIENCES	Anthropology	11
SOCIAL SCIENCES	Management	12
SOCIAL SCIENCES	Communication	13
SOCIAL SCIENCES	Operations Research & Management Science	14
SOCIAL SCIENCES	Law	15
SOCIAL SCIENCES	Social Sciences, Interdisciplinary	16
SOCIAL SCIENCES	International Relations	17
SOCIAL SCIENCES	Political Science	18
SOCIAL SCIENCES	Psychology, Multidisciplinary	19
SOCIAL SCIENCES	Psychology	20
SOCIAL SCIENCES	Economics	21
SOCIAL SCIENCES	Social Sciences, Mathematical Methods	22
SOCIAL SCIENCES	Area Studies	23
SOCIAL SCIENCES	Business, Finance	24
SOCIAL SCIENCES	Criminology & Penology	25
SOCIAL SCIENCES	Education, Scientific Disciplines	26
SOCIAL SCIENCES	Information Science & Library Science	27
SOCIAL SCIENCES	Social Issues	28
SOCIAL SCIENCES	Geography	29
SOCIAL SCIENCES	Social Work	30
SOCIAL SCIENCES	Environmental Studies	31
SOCIAL SCIENCES	Hospitality, Leisure, Sport & Tourism	32
SOCIAL SCIENCES	Public Administration	33
SOCIAL SCIENCES	Psychology, Applied	34
SOCIAL SCIENCES	Business	35
SOCIAL SCIENCES	Sociology	36
SOCIAL SCIENCES	Psychology, Biological	37
SOCIAL SCIENCES	Behavioral Sciences	38
SOCIAL SCIENCES	Demography	39

Área OECD	Disciplina	Identificador
SOCIAL SCIENCES	Planning & Development	40
SOCIAL SCIENCES	Psychology, Developmental	41
SOCIAL SCIENCES	Transportation	42
SOCIAL SCIENCES	Psychology, Educational	43
SOCIAL SCIENCES	Industrial Relations & Labor	44
SOCIAL SCIENCES	Psychology, Mathematical	45
SOCIAL SCIENCES	Education, Special	46

Tabla B.1: Asignación del identificador junto a su disciplina científica asociada de la figura B.1

Figura B.2: Identificador para cada una de las sub disciplinas dentro de la visualización.

Área OECD		Sub disciplina	Identificador
AGRICULTURAL SCIENCES		AGRICULTURE, FORESTRY, FISHERIES	1
AGRICULTURAL SCIENCES		ANIMAL AND DAIRY SCIENCE	2
AGRICULTURAL SCIENCES		VETERINARY SCIENCE	3
AGRICULTURAL SCIENCES		OTHER AGRICULTURAL SCIENCE	4
ENGINEERING TECHNOLOGY	AND	OTHER ENGINEERING AND TECHNOLOGIES	1
ENGINEERING TECHNOLOGY	AND	MEDICAL ENGINEERING	2
ENGINEERING TECHNOLOGY	AND	ENVIRONMENTAL BIOTECHNOLOGY	3
ENGINEERING TECHNOLOGY	AND	MECHANICAL ENGINEERING	4
ENGINEERING TECHNOLOGY	AND	CHEMICAL ENGINEERING	5
ENGINEERING TECHNOLOGY	AND	NANOTECHNOLOGY	6
ENGINEERING TECHNOLOGY	AND	ELECTRICAL ENG, ELECTRONIC ENG	7
ENGINEERING TECHNOLOGY	AND	INDUSTRIAL BIOTECHNOLOGY	8
ENGINEERING TECHNOLOGY	AND	CIVIL ENGINEERING	9
HUMANITIES		PHILOSOPHY, ETHICS AND RELIGION	1
HUMANITIES		HISTORY AND ARCHAEOLOGY	2
HUMANITIES		LANGUAGES AND LITERATURE	3
HUMANITIES		OTHER HUMANITIES	4
HUMANITIES		ART	5
MEDICAL AND HEALTH SCIENCES		CLINICAL MEDICINE	1
MEDICAL AND HEALTH SCIENCES		BASIC MEDICAL RESEARCH	2
MEDICAL AND HEALTH SCIENCES		HEALTH SCIENCES	3
NATURAL SCIENCES		CHEMICAL SCIENCES	1

Área OECD	Sub disciplina	Identificador
NATURAL SCIENCES	EARTH AND RELATED ENVIRONMENTAL SCIENCES	2
NATURAL SCIENCES	BIOLOGICAL SCIENCES	3
NATURAL SCIENCES	PHYSICAL SCIENCES AND ASTRONOMY	4
NATURAL SCIENCES	MATHEMATICS	5
NATURAL SCIENCES	COMPUTER AND INFORMATION SCIENCES	6
NATURAL SCIENCES	OTHER NATURAL SCIENCES	7
SOCIAL SCIENCES	SOCIOLOGY	1
SOCIAL SCIENCES	OTHER SOCIAL SCIENCES	2
SOCIAL SCIENCES	PSYCHOLOGY	3
SOCIAL SCIENCES	SOCIAL AND ECONOMIC GEOGRAPHY	4
SOCIAL SCIENCES	EDUCATIONAL SCIENCES	5
SOCIAL SCIENCES	ECONOMICS AND BUSINESS	6
SOCIAL SCIENCES	LAW	7
SOCIAL SCIENCES	POLITICAL SCIENCE	8
SOCIAL SCIENCES	MEDIA AND COMMUNICATION	9

Tabla B.2: Asignación del identificador junto a su sub disciplina científica asociada de la figura B.2

B.2. Argentina

Figura B.3: Red de complemento disciplinar para Argentina en el nivel disciplinar.

Figura B.4: Red de complemento disciplinar para Argentina en el nivel sub disciplinar.

Figura B.5: Red de complemento disciplinar para Argentina en el nivel de áreas de la OECD.

B.3. Brasil

Figura B.6: Red de complemento disciplinar para Brasil en el nivel disciplinar.

Figura B.7: Red de complemento disciplinar para Brasil en el nivel sub disciplinar.

Figura B.8: Red de complemento disciplinar para Brasil en el nivel de áreas de la OECD.

B.4. Chile

Figura B.9: Red de complemento disciplinar para Chile en el nivel disciplinar.

Figura B.10: Red de complemento disciplinar para Chile en el nivel sub disciplinar.

Figura B.11: Red de complemento disciplinar para Chile en el nivel de áreas de la OECD.

B.5. Colombia

Figura B.12: Red de complemento disciplinar para Colombia en el nivel disciplinar.

Figura B.13: Red de complemento disciplinar para Colombia en el nivel sub disciplinar.

Figura B.14: Red de complemento disciplinar para Colombia en el nivel de áreas de la OECD.

B.6. Cuba

Figura B.15: Red de complemento disciplinar para Colombia en el nivel disciplinar.

Figura B.16: Red de complemento disciplinar para Colombia en el nivel sub disciplinar.

Figura B.17: Red de complemento disciplinar para Colombia en el nivel de áreas de la OECD.

B.7. Finlandia

Figura B.18: Red de complemento disciplinar para Finlandia en el nivel disciplinar.

Figura B.19: Red de complemento disciplinar para Finlandia en el nivel sub disciplinar.

Figura B.20: Red de complemento disciplinar para Finlandia en el nivel de áreas de la OECD.

B.8. Grecia

Figura B.21: Red de complemento disciplinar para Grecia en el nivel disciplinar.

Figura B.22: Red de complemento disciplinar para Grecia en el nivel sub disciplinar.

Figura B.23: Red de complemento disciplinar para Grecia en el nivel de áreas de la OECD.

B.9. Islandia

Figura B.24: Red de complemento disciplinar para Islandia en el nivel disciplinar.

Figura B.25: Red de complemento disciplinar para Islandia en el nivel sub disciplinar.

Figura B.26: Red de complemento disciplinar para Islandia en el nivel de áreas de la OECD.

B.10. Israel

Figura B.27: Red de complemento disciplinar para Israel en el nivel disciplinar.

Figura B.28: Red de complemento disciplinar para Israel en el nivel sub disciplinar.

Figura B.29: Red de complemento disciplinar para Israel en el nivel de áreas de la OECD.

B.11. Nueva Zelandia

Figura B.30: Red de complemento disciplinar para Nueva Zelandia en el nivel disciplinar.

Figura B.31: Red de complemento disciplinar para Nueva Zelandia en el nivel sub disciplinar.

Figura B.32: Red de complemento disciplinar para Nueva Zelandia en el nivel de áreas de la OECD.

B.12. Polonia

Figura B.33: Red de complemento disciplinar para Polonia en el nivel disciplinar.

Figura B.34: Red de complemento disciplinar para Polonia en el nivel sub disciplinar.

Figura B.35: Red de complemento disciplinar para Polonia en el nivel de áreas de la OECD.

B.13. Portugal

Figura B.36: Red de complemento disciplinar para Portugal en el nivel disciplinar.

Figura B.37: Red de complemento disciplinar para Portugal en el nivel sub disciplinar.

Figura B.38: Red de complemento disciplinar para Portugal en el nivel de áreas de la OECD.

B.14. España

Figura B.39: Red de complemento disciplinar para España en el nivel disciplinar.

Figura B.40: Red de complemento disciplinar para España en el nivel sub disciplinar.

Figura B.41: Red de complemento disciplinar para España en el nivel de áreas de la OECD.

B.15. Venezuela

Figura B.42: Red de complemento disciplinar para Venezuela en el nivel disciplinar.

Figura B.43: Red de complemento disciplinar para Venezuela en el nivel sub disciplinar.

Figura B.44: Red de complemento disciplinar para Venezuela en el nivel de áreas de la OECD.

Apéndice C

Tabla de conversión disciplinar

Área OECD	Sub categoría		Disciplina
AGRICULTURAL SCIENCES	Agriculture, fisheries	forestry,	AGRICULTURE MULTIDISCIPLINARY
AGRICULTURAL SCIENCES	Agriculture, fisheries	forestry,	AGRICULTURE, MULTIDISCIPLINARY
AGRICULTURAL SCIENCES	Agriculture, fisheries	forestry,	AGRONOMY
AGRICULTURAL SCIENCES	Agriculture, fisheries	forestry,	FISHERIES
AGRICULTURAL SCIENCES	Agriculture, fisheries	forestry,	FORESTRY
AGRICULTURAL SCIENCES	Agriculture, fisheries	forestry,	HORTICULTURE
AGRICULTURAL SCIENCES	Agriculture, fisheries	forestry,	SOIL SCIENCE
AGRICULTURAL SCIENCES	Animal and dairy science		AGRICULTURE, DAIRY & ANIMAL SCIENCE
AGRICULTURAL SCIENCES	Other science	agricultural	AGRICULTURAL ECONOMICS & POLICY
AGRICULTURAL SCIENCES	Other science	agricultural	AGRICULTURAL ENGINEERING
AGRICULTURAL SCIENCES	Other science	agricultural	AGRICULTURE ENGINEERING
AGRICULTURAL SCIENCES	Other science	agricultural	FOOD SCIENCE & TECHNOLOGY

Área OECD	Sub categoría	Disciplina
AGRICULTURAL SCIENCES	Veterinary science	VETERINARY SCIENCES
ENGINEERING AND TECHNOLOGY	Chemical engineering	ENGINEERING, CHEMICAL
ENGINEERING AND TECHNOLOGY	Civil engineering	CONSTRUCTION & BUILDING TECHNOLOGY
ENGINEERING AND TECHNOLOGY	Civil engineering	ENGINEERING, CIVIL
ENGINEERING AND TECHNOLOGY	Civil engineering	TRANSPORTATION SCIENCE & TECHNOLOGY
ENGINEERING AND TECHNOLOGY	Electrical eng, electronic eng	AUTOMATION & CONTROL SYSTEM
ENGINEERING AND TECHNOLOGY	Electrical eng, electronic eng	AUTOMATION & CONTROL SYSTEMS
ENGINEERING AND TECHNOLOGY	Electrical eng, electronic eng	COMPUTER SCIENCE, HARDWARE & ARCHITECTURE
ENGINEERING AND TECHNOLOGY	Electrical eng, electronic eng	ENGINEERING, ELECTRICAL & ELECTRONIC
ENGINEERING AND TECHNOLOGY	Electrical eng, electronic eng	ROBOTICS
ENGINEERING AND TECHNOLOGY	Electrical eng, electronic eng	TELECOMMUNICATIONS
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	BIOTECHNOLOGY & APPLIED MICROBIOLOGY
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	ENERGY & FUELS
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	ENGINEERING, ENVIRONMENTAL
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	ENGINEERING, GEOLOGICAL
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	ENGINEERING, MARINE
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	ENGINEERING, OCEAN
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	ENGINEERING, PETROLEUM

Área OECD	Sub categoría	Disciplina
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	MINING & MINERAL PROCESSING
ENGINEERING AND TECHNOLOGY	Environmental biotechnology	REMOTE SENSING
ENGINEERING AND TECHNOLOGY	Industrial biotechnology	MATERIALS SCIENCE, BIOMATERIALS
ENGINEERING AND TECHNOLOGY	Mechanical engineering	ENGINEERING, AEROSPACE
ENGINEERING AND TECHNOLOGY	Mechanical engineering	ENGINEERING, MECHANICAL
ENGINEERING AND TECHNOLOGY	Mechanical engineering	MECHANICS
ENGINEERING AND TECHNOLOGY	Mechanical engineering	NUCLEAR SCIENCE & TECHNOLOGY
ENGINEERING AND TECHNOLOGY	Mechanical engineering	THERMODYNAMICS
ENGINEERING AND TECHNOLOGY	Medical engineering	CELL & TISSUE ENGINEERING
ENGINEERING AND TECHNOLOGY	Medical engineering	ENGINEERING, BIOMEDICAL
ENGINEERING AND TECHNOLOGY	Medical engineering	MATERIALS SCIENCE, CERAMICS
ENGINEERING AND TECHNOLOGY	Medical engineering	MATERIALS SCIENCE, CHARACTERIZATION & TESTING
ENGINEERING AND TECHNOLOGY	Medical engineering	MATERIALS SCIENCE, COATINGS & FILMS
ENGINEERING AND TECHNOLOGY	Medical engineering	MATERIALS SCIENCE, COMPOSITES
ENGINEERING AND TECHNOLOGY	Medical engineering	MATERIALS SCIENCE, MULTIDISCIPLINARY
ENGINEERING AND TECHNOLOGY	Medical engineering	MATERIALS SCIENCE, PAPER & WOOD
ENGINEERING AND TECHNOLOGY	Medical engineering	MATERIALS SCIENCE, TEXTILES
ENGINEERING AND TECHNOLOGY	Medical engineering	MEDICAL LABORATORY TECHNOLOGY

Área OECD	Sub categoría	Disciplina
ENGINEERING AND TECHNOLOGY	Medical engineering	METALLURGY & METALLURGICAL ENGINEERING
ENGINEERING AND TECHNOLOGY	Nanotechnology	NANOSCIENCE & NANOTECHNOLOGY
ENGINEERING AND TECHNOLOGY	Other engineering and technologies	ENGINEERING, INDUSTRIAL
ENGINEERING AND TECHNOLOGY	Other engineering and technologies	ENGINEERING, MANUFACTURING
ENGINEERING AND TECHNOLOGY	Other engineering and technologies	ENGINEERING, MULTIDISCIPLINARY
ENGINEERING AND TECHNOLOGY	Other engineering and technologies	IMAGING SCIENCE & PHOTOGRAPHIC TECHNOLOGY
ENGINEERING AND TECHNOLOGY	Other engineering and technologies	INSTRUMENTS & INSTRUMENTATION
ENGINEERING AND TECHNOLOGY	Other engineering and technologies	MICROSCOPY
ENGINEERING AND TECHNOLOGY	Other engineering and technologies	SPECTROSCOPY
ENGINEERING AND TECHNOLOGY	Other engineering and technologies	FOOD SCIENCE & TECHNOLOGY
ENGINEERING AND TECHNOLOGY	Environmental engineering	GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY
HUMANITIES	Art	ARCHITECTURE
HUMANITIES	Art	ART
HUMANITIES	Art	DANCE
HUMANITIES	Art	FILM, RADIO, TELEVISION
HUMANITIES	Art	MUSIC
HUMANITIES	Art	THEATER
HUMANITIES	History and archaeology	ARCHAEOLOGY
HUMANITIES	History and archaeology	HISTORY
HUMANITIES	History and archaeology	HISTORY & PHILOSOPHY OF SCIENCE
HUMANITIES	History and archaeology	HISTORY OF SOCIAL SCIENCES
HUMANITIES	History and archaeology	MEDIEVAL & RENAISSANCE STUDIES

Área OECD	Sub categoría	Disciplina
HUMANITIES	Languages and literature	CLASSICS
HUMANITIES	Languages and literature	ENGLISH
HUMANITIES	Languages and literature	FOLKLORE
HUMANITIES	Languages and literature	LANGUAGE & LINGUISTICS
HUMANITIES	Languages and literature	LINGUISTICS
HUMANITIES	Languages and literature	LITERARY REVIEWS
HUMANITIES	Languages and literature	LITERARY THEORY & CRITICISM
HUMANITIES	Languages and literature	LITERATURE
HUMANITIES	Languages and literature	LITERATURE, AFRICAN, AUSTRALIAN, CANADIAN
HUMANITIES	Languages and literature	LITERATURE, AMERICAN
HUMANITIES	Languages and literature	LITERATURE, BRITISH ISLES
HUMANITIES	Languages and literature	LITERATURE, GERMAN, DUTCH, SCANDINAVIAN
HUMANITIES	Languages and literature	LITERATURE, ROMANCE
HUMANITIES	Languages and literature	LITERATURE, SLAVIC
HUMANITIES	Languages and literature	POETRY
HUMANITIES	Languages and literature	SPANISH
HUMANITIES	Other Humanities	HUMANITIES, MULTIDISCIPLINARY
HUMANITIES	Philosophy, ethics and religion	ETHICS
HUMANITIES	Philosophy, ethics and religion	PHILOSOPHY
HUMANITIES	Philosophy, ethics and religion	RELIGION

Área OECD	Sub categoría	Disciplina
MEDICAL AND HEALTH SCIENCES	Basic medical research	ANATOMY & MORPHOLOGY
MEDICAL AND HEALTH SCIENCES	Basic medical research	CHEMISTRY, MEDICAL
MEDICAL AND HEALTH SCIENCES	Basic medical research	IMMUNOLOGY
MEDICAL AND HEALTH SCIENCES	Basic medical research	MEDICINE, RESEARCH & EXPERIMENTAL
MEDICAL AND HEALTH SCIENCES	Basic medical research	NEUROSCIENCES
MEDICAL AND HEALTH SCIENCES	Basic medical research	PATHOLOGY
MEDICAL AND HEALTH SCIENCES	Basic medical research	PHARMACOLOGY & PHARMACY
MEDICAL AND HEALTH SCIENCES	Basic medical research	PHYSIOLOGY
MEDICAL AND HEALTH SCIENCES	Basic medical research	PSYCHOLOGY, CLINICAL
MEDICAL AND HEALTH SCIENCES	Basic medical research	TOXICOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	ALLERGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	ANDROLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	ANESTHESIOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	AUDIOLOGY & SPEECH-LANGUAGE PATHOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	AUDIOLOGY & SPEECH-LENGUAJE PATHOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	CARDIAC & CARDIOVASCULAR SYSTEMS
MEDICAL AND HEALTH SCIENCES	Clinical medicine	CLINICAL NEUROLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	CRITICAL CARE MEDICINE

Área OECD	Sub categoría	Disciplina
MEDICAL AND HEALTH SCIENCES	Clinical medicine	DENTISTRY, ORAL SURGERY & MEDICINE
MEDICAL AND HEALTH SCIENCES	Clinical medicine	DERMATOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	EMERGENCY MEDICINE
MEDICAL AND HEALTH SCIENCES	Clinical medicine	ENDOCRINOLOGY & METABOLISM
MEDICAL AND HEALTH SCIENCES	Clinical medicine	GASTROENTEROLOGY & HEPATOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	GERIATRICS & GERONTOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	GERONTOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	HEMATOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	INTEGRATIVE & COMPLEMENTARY MEDICINE
MEDICAL AND HEALTH SCIENCES	Clinical medicine	MEDICINE, GENERAL & INTERNAL
MEDICAL AND HEALTH SCIENCES	Clinical medicine	NEUROIMAGING
MEDICAL AND HEALTH SCIENCES	Clinical medicine	OBSTETRICS & GYNECOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	ONCOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	OPHTHALMOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	ORTHOPEDICS
MEDICAL AND HEALTH SCIENCES	Clinical medicine	OTORHINOLARYNGOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	PEDIATRICS
MEDICAL AND HEALTH SCIENCES	Clinical medicine	PERIPHERAL VASCULAR DISEASE
MEDICAL AND HEALTH SCIENCES	Clinical medicine	PSYCHIATRY

Área OECD	Sub categoría	Disciplina
MEDICAL AND HEALTH SCIENCES	Clinical medicine	RADIOLOGY, NUCLEAR MEDICINE & MEDICAL IMAGING
MEDICAL AND HEALTH SCIENCES	Clinical medicine	RESPIRATORY SYSTEM
MEDICAL AND HEALTH SCIENCES	Clinical medicine	RHEUMATOLOGY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	SURGERY
MEDICAL AND HEALTH SCIENCES	Clinical medicine	TRANSPLANTATION
MEDICAL AND HEALTH SCIENCES	Clinical medicine	UROLOGY & NEPHROLOGY
MEDICAL AND HEALTH SCIENCES	Health sciences	HEALTH CARE SCIENCES & SERVICES
MEDICAL AND HEALTH SCIENCES	Health sciences	HEALTH POLICY & SERVICES
MEDICAL AND HEALTH SCIENCES	Health sciences	INFECTIOUS DISEASES
MEDICAL AND HEALTH SCIENCES	Health sciences	MEDICAL ETHICS
MEDICAL AND HEALTH SCIENCES	Health sciences	MEDICAL INFORMATICS
MEDICAL AND HEALTH SCIENCES	Health sciences	MEDICINE, LEGAL
MEDICAL AND HEALTH SCIENCES	Health sciences	NURSING
MEDICAL AND HEALTH SCIENCES	Health sciences	NUTRITION & DIETETICS
MEDICAL AND HEALTH SCIENCES	Health sciences	PARASITOLOGY
MEDICAL AND HEALTH SCIENCES	Health sciences	PRIMARY HEALTH CARE
MEDICAL AND HEALTH SCIENCES	Health sciences	PSYCHOLOGY, PSYCHOANALYSIS
MEDICAL AND HEALTH SCIENCES	Health sciences	PUBLIC, ENVIRONMENTAL & OCCUPATIONAL HEALTH
MEDICAL AND HEALTH SCIENCES	Health sciences	REHABILITATION

Área OECD	Sub categoría	Disciplina
MEDICAL AND HEALTH SCIENCES	Health sciences	SOCIAL SCIENCES, BIOMEDICAL
MEDICAL AND HEALTH SCIENCES	Health sciences	SPORT SCIENCES
MEDICAL AND HEALTH SCIENCES	Health sciences	SUBSTANCE ABUSE
MEDICAL AND HEALTH SCIENCES	Health sciences	TROPICAL MEDICINE
NATURAL SCIENCES	Biological sciences	BIOCHEMICAL RESEARCH METHODS
NATURAL SCIENCES	Biological sciences	BIOCHEMISTRY & MOLECULAR BIOLOGY
NATURAL SCIENCES	Biological sciences	BIODIVERSITY CONSERVATION
NATURAL SCIENCES	Biological sciences	BIOLOGY
NATURAL SCIENCES	Biological sciences	BIOPHYSICS
NATURAL SCIENCES	Biological sciences	CELL BIOLOGY
NATURAL SCIENCES	Biological sciences	DEVELOPMENTAL BIOLOGY
NATURAL SCIENCES	Biological sciences	ECOLOGY
NATURAL SCIENCES	Biological sciences	ENTOMOLOGY
NATURAL SCIENCES	Biological sciences	EVOLUTIONARY BIOLOGY
NATURAL SCIENCES	Biological sciences	GENETICS & HEREDITY
NATURAL SCIENCES	Biological sciences	LIMNOLOGY
NATURAL SCIENCES	Biological sciences	MARINE & FRESHWATER BIOLOGY
NATURAL SCIENCES	Biological sciences	MATHEMATICAL & COMPUTATIONAL BIOLOGY
NATURAL SCIENCES	Biological sciences	MICROBIOLOGY
NATURAL SCIENCES	Biological sciences	MYCOLOGY

Área OECD		Sub categoría	Disciplina
NATURAL CES	SCIEN-	Biological sciences	ORNITHOLOGY
NATURAL CES	SCIEN-	Biological sciences	PLANT SCIENCES
NATURAL CES	SCIEN-	Biological sciences	REPRODUCTIVE BIOLOGY
NATURAL CES	SCIEN-	Biological sciences	VIROLOGY
NATURAL CES	SCIEN-	Biological sciences	ZOOLOGY
NATURAL CES	SCIEN-	Chemical sciences	CHEMISTRY
NATURAL CES	SCIEN-	Chemical sciences	CHEMISTRY, ANALYTICAL
NATURAL CES	SCIEN-	Chemical sciences	CHEMISTRY, AP- PLIED
NATURAL CES	SCIEN-	Chemical sciences	CHEMISTRY, INOR- GANIC & NUCLEAR
NATURAL CES	SCIEN-	Chemical sciences	CHEMISTRY, MULTI- DISCIPLINARY
NATURAL CES	SCIEN-	Chemical sciences	CHEMISTRY, ORGA- NIC
NATURAL CES	SCIEN-	Chemical sciences	CHEMISTRY, PHYSI- CAL
NATURAL CES	SCIEN-	Chemical sciences	CRYSTALLOGRAPHY
NATURAL CES	SCIEN-	Chemical sciences	ELECTROCHEMISTRY
NATURAL CES	SCIEN-	Chemical sciences	POLYMER SCIENCE
NATURAL CES	SCIEN-	Computer and informa- tion sciences	COMPUTER SCIEN- CE, ARTIFICIAL INTELLIGENCE
NATURAL CES	SCIEN-	Computer and informa- tion sciences	COMPUTER SCIEN- CE, CYBERNETICS
NATURAL CES	SCIEN-	Computer and informa- tion sciences	COMPUTER SCIEN- CE, INFORMATION SYSTEMS
NATURAL CES	SCIEN-	Computer and informa- tion sciences	COMPUTER SCIEN- CE, INTERDISCI- PLINARY APPLICA- TIONS

Área OECD		Sub categoría	Disciplina
NATURAL CES	SCIEN-	Computer and informa- tion sciences	COMPUTER SCIEN- CE, SOFTWARE ENGINEERING
NATURAL CES	SCIEN-	Computer and informa- tion sciences	COMPUTER SCIEN- CE, THEORY & METHODS
NATURAL CES	SCIEN-	Computer and informa- tion sciences	LOGIC
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	ENVIRONMENTAL SCIENCES
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	GEOCHEMISTRY & GEOPHYSICS
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	GEOGRAPHY, PHY- SICAL
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	GEOLOGY
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	GEOSCIENCES, MUL- TIDISCIPLINARY
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	METEOROLOGY & ATMOSPHERIC SCIENCES
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	MINERALOGY
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	OCEANOGRAPHY
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	PALEONTOLOGY
NATURAL CES	SCIEN-	Earth and related envi- ronmental sciences	WATER RESOURCES
NATURAL CES	SCIEN-	Mathematics	MATHEMATICS
NATURAL CES	SCIEN-	Mathematics	MATHEMATICS, AP- PLIED
NATURAL CES	SCIEN-	Mathematics	MATHEMATICS, IN- TERDISCIPLINARY APPLICATIONS
NATURAL CES	SCIEN-	Mathematics	PHYSICS, MATHE- MATICAL
NATURAL CES	SCIEN-	Mathematics	STATISTICS & PRO- BABILITY
NATURAL CES	SCIEN-	Other natural sciences	MULTIDISCIPLINARY SCIENCES

Área OECD		Sub categoría	Disciplina
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	ACOUSTICS
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	ASTRONOMY & ASTROPHYSICS
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	OPTICS
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	PHYSICS, APPLIED
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	PHYSICS, ATOMIC, MOLECULAR & CHEMICAL
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	PHYSICS, CONDENSED MATTER
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	PHYSICS, FLUIDS & PLASMAS
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	PHYSICS, MULTIDISCIPLINARY
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	PHYSICS, NUCLEAR
NATURAL SCIENCES	SCIENCES	Physical sciences and astronomy	PHYSICS, PARTICLES & FIELDS
NATURAL SCIENCES	SCIENCES	Mathematics	LOGIC
NATURAL SCIENCES	SCIENCES	Chemical sciences	GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY
NATURAL SCIENCES	SCIENCES	Earth and related environmental sciences	GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY
SOCIAL SCIENCES		Economics and business	BUSINESS
SOCIAL SCIENCES		Economics and business	BUSINESS, FINANCE
SOCIAL SCIENCES		Economics and business	ECONOMICS
SOCIAL SCIENCES		Educational sciences	EDUCATION & EDUCATIONAL RESEARCH
SOCIAL SCIENCES		Educational sciences	EDUCATION, SCIENTIFIC DISCIPLINES
SOCIAL SCIENCES		Educational sciences	EDUCATION, SPECIAL
SOCIAL SCIENCES		Economics and business	INDUSTRIAL RELATIONS & LABOR
SOCIAL SCIENCES		Economics and business	MANAGEMENT

Área OECD	Sub categoría	Disciplina
SOCIAL SCIENCES	Economics and business	OPERATIONS RE- SEARCH & MANA- GEMENT SCIENCE
SOCIAL SCIENCES	Law	CRIMINOLOGY & PENOLOGY
SOCIAL SCIENCES	Law	LAW
SOCIAL SCIENCES	Media and communica- tion	COMMUNICATION
SOCIAL SCIENCES	Media and communica- tion	INFORMATION SCIENCE & LIBRARY SCIENCE
SOCIAL SCIENCES	Other social sciences	ASIAN STUDIES
SOCIAL SCIENCES	Other social sciences	CULTURAL STUDIES
SOCIAL SCIENCES	Other social sciences	HOSPITALITY, LEI- SURE, SPORT & TOURISM
SOCIAL SCIENCES	Other social sciences	SOCIAL SCIENCES, INTERDISCIPLI- NARY
SOCIAL SCIENCES	Political science	INTERNATIONAL RELATIONS
SOCIAL SCIENCES	Political science	POLITICAL SCIENCE
SOCIAL SCIENCES	Political science	PUBLIC ADMINIS- TRATION
SOCIAL SCIENCES	Psychology	BEHAVIORAL SCIEN- CES
SOCIAL SCIENCES	Psychology	ERGONOMICS
SOCIAL SCIENCES	Psychology	PSYCHOLOGY
SOCIAL SCIENCES	Psychology	PSYCHOLOGY, AP- PLIED
SOCIAL SCIENCES	Psychology	PSYCHOLOGY, BIO- LOGICAL
SOCIAL SCIENCES	Psychology	PSYCHOLOGY, DE- VELOPMENTAL
SOCIAL SCIENCES	Psychology	PSYCHOLOGY, EDU- CATIONAL
SOCIAL SCIENCES	Psychology	PSYCHOLOGY, EX- PERIMENTAL
SOCIAL SCIENCES	Psychology	PSYCHOLOGY, MAT- HEMATICAL
SOCIAL SCIENCES	Psychology	PSYCHOLOGY, MUL- TIDISCIPLINARY

Área OECD	Sub categoría	Disciplina
SOCIAL SCIENCES	Psychology	PSYCHOLOGY, SOCIAL
SOCIAL SCIENCES	Social and economic geography	AREA STUDIES
SOCIAL SCIENCES	Social and economic geography	ENVIRONMENTAL STUDIES
SOCIAL SCIENCES	Social and economic geography	GEOGRAPHY
SOCIAL SCIENCES	Social and economic geography	PLANNING & DEVELOPMENT
SOCIAL SCIENCES	Social and economic geography	TRANSPORTATION
SOCIAL SCIENCES	Social and economic geography	URBAN STUDIES
SOCIAL SCIENCES	Sociology	ANTHROPOLOGY
SOCIAL SCIENCES	Sociology	DEMOGRAPHY
SOCIAL SCIENCES	Sociology	ETHNIC STUDIES
SOCIAL SCIENCES	Sociology	FAMILY STUDIES
SOCIAL SCIENCES	Sociology	SOCIAL ISSUES

Tabla C.1: Tabla de conversión disciplinar

Apéndice D

Ponderación de importancia disciplinar por país

En este apéndice se presenta en las siguientes tablas cómo fue desarrollado el Índice de Complejidad Disciplinar (ICD) de cada uno de los países tratados en el capítulo 5, junto con la ponderación de importancia disciplinar para cada una de las disciplinas del estudio.

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Acoustics	0.002	0.002	0.002	0.002
Agricultural Economics & Policy	0.000	0.000	0.000	0.000
Agricultural Engineering	0.000	0.000	0.000	0.000
Agriculture, Dairy & Animal Science	0.000	0.000	0.000	0.000
Agriculture, Multidisciplinary	0.005	0.003	0.004	0.005
Agronomy	0.022	0.010	0.019	0.022
Allergy	0.000	0.000	0.000	0.000
Anatomy & Morphology	0.000	0.000	0.000	0.000
Andrology	0.000	0.000	0.000	0.000
Anesthesiology	0.000	0.000	0.000	0.000
Anthropology	0.001	0.002	0.001	0.001
Archaeology	0.000	0.000	0.000	0.000
Architecture	0.000	0.000	0.000	0.000
Area Studies	0.000	0.000	0.000	0.000
Art	0.000	0.000	0.000	0.000
Asian Studies	0.000	0.000	0.000	0.000
Astronomy & Astrophysics	0.139	0.138	0.181	0.239

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Audiology & Speech-Language Pathology	0.000	0.000	0.000	0.000
Automation & Control Systems	0.004	0.004	0.010	0.003
Behavioral Sciences	0.001	0.001	0.001	0.001
Biochemical Research Methods	0.025	0.026	0.038	0.025
Biochemistry & Molecular Biology	20.680	16.662	23.966	21.574
Biodiversity Conservation	0.001	0.001	0.001	0.001
Biology	0.056	0.032	0.055	0.045
Biophysics	0.006	0.006	0.007	0.005
Biotechnology & Applied Microbiology	1.714	1.773	1.793	2.017
Business	0.001	0.001	0.001	0.001
Business, Finance	0.000	0.000	0.000	0.000
Cardiac & Cardiovascular Systems	0.022	0.015	0.029	0.028
Cell & Tissue Engineering	0.000	0.000	0.000	0.000
Cell Biology	0.511	0.453	0.682	0.508
Chemistry, Analytical	0.278	0.269	0.253	0.197
Chemistry, Applied	0.080	0.077	0.089	0.078
Chemistry, Inorganic & Nuclear	0.008	0.007	0.008	0.005
Chemistry, Medicinal	0.010	0.009	0.010	0.010
Chemistry, Multidisciplinary	1.779	1.191	1.056	0.777
Chemistry, Organic	0.001	0.001	0.001	0.001
Chemistry, Physical	9.728	10.533	9.673	10.160
Classics	0.000	0.000	0.000	0.000
Clinical Neurology	0.212	0.142	0.134	0.115
Communication	0.000	0.000	0.001	0.000
Computer Science, Artificial Intelligence	0.049	0.047	0.051	0.062
Computer Science, Cybernetics	0.000	0.000	0.000	0.000
Computer Science, Hardware & Architecture	0.000	0.000	0.000	0.000
Computer Science, Information Systems	0.005	0.006	0.006	0.003

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Computer Science, Interdisciplinary Applications	0.914	0.754	1.032	0.804
Computer Science, Software Engineering	0.000	0.000	0.000	0.001
Computer Science, Theory & Methods	0.003	0.003	0.003	0.002
Construction & Building Technology	0.001	0.001	0.001	0.002
Criminology & Penology	0.000	0.000	0.000	0.000
Critical Care Medicine	0.000	0.000	0.000	0.000
Crystallography	0.000	0.000	0.000	0.000
Cultural Studies	0.000	0.000	0.000	0.000
Demography	0.000	0.000	0.000	0.000
Dentistry, Oral Surgery & Medicine	0.010	0.012	0.011	0.011
Dermatology	0.001	0.001	0.001	0.001
Developmental Biology	0.000	0.000	0.000	0.000
Ecology	1.624	1.580	1.447	2.023
Economics	0.315	0.304	0.285	0.264
Education & Educational Research	0.024	0.011	0.048	0.010
Education, Scientific Disciplines	0.000	0.000	0.000	0.000
Education, Special	0.000	0.000	0.000	0.000
Electrochemistry	0.001	0.001	0.001	0.001
Emergency Medicine	0.000	0.000	0.000	0.000
Endocrinology & Metabolism	0.012	0.010	0.011	0.010
Energy & Fuels	0.946	1.028	0.790	0.873
Engineering, Aerospace	0.000	0.000	0.000	0.000
Engineering, Biomedical	0.037	0.034	0.032	0.025
Engineering, Chemical	0.787	0.833	0.795	0.930
Engineering, Civil	0.112	0.104	0.100	0.113
Engineering, Electrical & Electronic	3.084	3.018	8.137	3.012
Engineering, Environmental	0.003	0.003	0.002	0.003

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Engineering, Geological	0.000	0.000	0.000	0.000
Engineering, Industrial	0.000	0.000	0.000	0.000
Engineering, Manufacturing	0.001	0.000	0.000	0.000
Engineering, Marine	0.000	0.000	0.000	0.000
Engineering, Mechanical	0.022	0.019	0.027	0.023
Engineering, Multidisciplinary	0.003	0.003	0.003	0.001
Engineering, Ocean	0.000	0.000	0.000	0.000
Engineering, Petroleum	0.000	0.000	0.000	0.000
Entomology	0.001	0.001	0.000	0.001
Environmental Sciences	15.947	14.802	13.206	17.844
Environmental Studies	0.056	0.038	0.042	0.052
Ergonomics	0.000	0.000	0.000	0.000
Ethics	0.000	0.000	0.000	0.000
Ethnic Studies	0.000	0.000	0.000	0.000
Evolutionary Biology	0.000	0.000	0.000	0.000
Family Studies	0.000	0.000	0.000	0.000
Film, Radio, Television	0.000	0.000	0.000	0.000
Fisheries	0.003	0.003	0.003	0.003
Folklore	0.000	0.000	0.000	0.000
Food Science & Technology	0.437	0.345	0.439	0.365
Forestry	0.007	0.002	0.004	0.005
Gastroenterology & Hepatology	0.001	0.000	0.001	0.001
Genetics & Heredity	0.621	0.452	0.735	0.689
Geochemistry & Geophysics	0.002	0.002	0.003	0.002
Geography	0.000	0.000	0.000	0.000
Geography, Physical	0.010	0.008	0.011	0.011
Geology	0.000	0.000	0.000	0.000
Geosciences, Multidisciplinary	0.508	0.527	0.692	0.566
Geriatrics & Gerontology	0.002	0.002	0.002	0.002
Gerontology	0.000	0.000	0.000	0.000

DISCIPLINA	Argentina	Brasil	Chile	Colombia
GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY	0.000	0.000	0.000	0.000
Health Care Sciences & Services	0.002	0.001	0.001	0.001
Health Policy & Services	0.000	0.000	0.000	0.000
Hematology	0.001	0.001	0.001	0.001
History	0.001	0.000	0.001	0.000
History & Philosophy Of Science	0.000	0.000	0.000	0.000
History Of Social Sciences	0.000	0.000	0.000	0.000
Horticulture	0.000	0.000	0.000	0.000
Hospitality, Leisure, Sport & Tourism	0.000	0.000	0.000	0.000
Humanities, Multidisciplinary	0.000	0.000	0.000	0.000
Imaging Science & Photographic Technology	0.000	0.000	0.001	0.000
Immunology	1.117	0.939	0.982	1.613
Industrial Relations & Labor	0.000	0.000	0.000	0.000
Infectious Diseases	0.085	0.067	0.041	0.086
Information Science & Library Science	0.000	0.000	0.000	0.000
Instruments & Instrumentation	0.030	0.031	0.106	0.037
Integrative & Complementary Medicine	0.000	0.000	0.000	0.000
International Relations	0.000	0.000	0.000	0.000
Language & Linguistics	0.000	0.000	0.000	0.000
Law	0.000	0.000	0.000	0.000
Limnology	0.000	0.000	0.000	0.000
Linguistics	0.000	0.000	0.000	0.000
Literary Reviews	0.000	0.000	0.000	0.000
Literary Theory & Criticism	0.000	0.000	0.000	0.000
Literature	0.000	0.000	0.000	0.000

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Literature, African, Australian, Canadian	0.000	0.000	0.000	0.000
Literature, American	0.000	0.000	0.000	0.000
Literature, British Isles	0.000	0.000	0.000	0.000
Literature, German, Dutch, Scandinavian	0.000	0.000	0.000	0.000
Literature, Romance	0.000	0.000	0.000	0.000
Literature, Slavic	0.000	0.000	0.000	0.000
Logic	0.000	0.000	0.000	0.000
Management	0.011	0.007	0.010	0.007
Marine & Freshwater Biology	0.033	0.034	0.032	0.028
Materials Science, Biomaterials	0.000	0.000	0.000	0.000
Materials Science, Ceramics	0.000	0.000	0.000	0.000
Materials Science, Characterization & Testing	0.000	0.000	0.000	0.000
Materials Science, Coatings & Films	0.000	0.000	0.000	0.000
Materials Science, Composites	0.000	0.000	0.000	0.000
Materials Science, Multidisciplinary	8.500	7.459	6.922	7.115
Materials Science, Paper & Wood	0.000	0.000	0.000	0.000
Materials Science, Textiles	0.000	0.000	0.000	0.000
Mathematical & Computational Biology	0.002	0.001	0.001	0.001
Mathematics	0.000	0.000	0.000	0.000
Mathematics, Applied	0.035	0.041	0.044	0.028
Mathematics, Interdisciplinary Applications	0.011	0.009	0.010	0.007
Mechanics	0.039	0.030	0.037	0.029
Medical Ethics	0.000	0.000	0.000	0.000
Medical Informatics	0.000	0.000	0.000	0.000
Medical Laboratory Technology	0.000	0.000	0.000	0.000

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Medicine, General & Internal	1.000	0.471	0.283	0.503
Medicine, Legal	0.000	0.000	0.000	0.000
Medicine, Research & Experimental	0.048	0.043	0.051	0.069
Medieval & Renaissance Studies	0.000	0.000	0.000	0.000
Metallurgy & Metallurgical Engineering	0.000	0.000	0.000	0.000
Meteorology & Atmospheric Sciences	0.011	0.012	0.016	0.016
Microbiology	0.082	0.084	0.094	0.138
Microscopy	0.000	0.000	0.000	0.000
Mineralogy	0.000	0.000	0.000	0.000
Mining & Mineral Processing	0.000	0.000	0.000	0.000
Multidisciplinary Sciences	0.609	0.402	0.788	0.766
Music	0.000	0.000	0.000	0.000
Mycology	0.000	0.000	0.000	0.001
Nanoscience & Nanotechnology	0.076	0.068	0.063	0.048
Neuroimaging	0.000	0.000	0.000	0.000
Neurosciences	4.177	3.170	3.885	4.070
Nuclear Science & Technology	0.010	0.008	0.008	0.014
Nursing	0.001	0.001	0.001	0.000
Nutrition & Dietetics	0.070	0.058	0.055	0.060
Obstetrics & Gynecology	0.024	0.017	0.035	0.020
Oceanography	0.008	0.006	0.006	0.005
Oncology	1.321	1.134	1.311	1.266
Operations Research & Management Science	0.014	0.021	0.017	0.017
Ophthalmology	0.000	0.000	0.000	0.000
Optics	0.022	0.023	0.017	0.020
Ornithology	0.000	0.000	0.000	0.000
Orthopedics	0.000	0.000	0.000	0.000
Otorhinolaryngology	0.000	0.000	0.000	0.000
Paleontology	0.000	0.000	0.000	0.000
Parasitology	0.001	0.000	0.000	0.001
Pathology	0.003	0.003	0.004	0.002
Pediatrics	0.051	0.062	0.073	0.060

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Peripheral Vascular Disease	0.003	0.003	0.003	0.002
Pharmacology & Pharmacy	1.815	1.829	1.838	1.773
Philosophy	0.000	0.000	0.000	0.000
Physics, Applied	0.076	0.075	0.072	0.049
Physics, Atomic, Molecular & Chemical	0.013	0.013	0.013	0.014
Physics, Condensed Matter	0.007	0.007	0.007	0.005
Physics, Fluids & Plasmas	0.000	0.000	0.000	0.000
Physics, Mathematical	0.005	0.005	0.005	0.005
Physics, Multidisciplinary	0.100	0.090	0.119	0.127
Physics, Nuclear	0.010	0.008	0.010	0.018
Physics, Particles & Fields	0.048	0.036	0.040	0.057
Physiology	0.007	0.005	0.008	0.007
Planning & Development	0.000	0.000	0.000	0.000
Plant Sciences	0.805	0.599	0.772	0.653
Poetry	0.000	0.000	0.000	0.000
Political Science	0.000	0.001	0.000	0.000
Polymer Science	0.053	0.044	0.041	0.045
Primary Health Care	0.000	0.000	0.000	0.000
Psychiatry	0.442	0.401	0.313	1.014
Psychology	0.004	0.005	0.003	0.005
Psychology, Applied	0.000	0.000	0.000	0.000
Psychology, Biological	0.000	0.000	0.000	0.000
Psychology, Clinical	0.000	0.000	0.000	0.000
Psychology, Developmental	0.000	0.001	0.001	0.001
Psychology, Educational	0.000	0.000	0.000	0.000
Psychology, Experimental	0.005	0.004	0.006	0.004
Psychology, Mathematical	0.000	0.000	0.000	0.000
Psychology, Multidisciplinary	0.001	0.001	0.001	0.001
Psychology, Psychoanalysis	0.000	0.000	0.000	0.000

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Psychology, Social	0.001	0.000	0.001	0.001
Public Administration	0.000	0.000	0.000	0.000
Public, Environmental & Occupational Health	2.647	1.606	1.941	2.135
Radiology, Nuclear Medicine & Medical Imaging	0.117	0.101	0.124	0.093
Rehabilitation	0.004	0.003	0.002	0.003
Religion	0.000	0.000	0.000	0.000
Remote Sensing	0.000	0.000	0.000	0.000
Reproductive Biology	0.000	0.000	0.001	0.000
Respiratory System	0.001	0.000	0.001	0.001
Rheumatology	0.000	0.000	0.000	0.000
Robotics	0.000	0.000	0.000	0.000
Social Issues	0.000	0.000	0.000	0.000
Social Sciences, Bio- medical	0.000	0.000	0.000	0.000
Social Sciences, Inter- disciplinary	0.000	0.000	0.001	0.000
Social Sciences, Mat- hematical Methods	0.000	0.000	0.000	0.000
Social Work	0.000	0.000	0.000	0.000
Sociology	0.000	0.001	0.001	0.001
Soil Science	0.000	0.000	0.000	0.000
Spectroscopy	0.000	0.000	0.000	0.000
Sport Sciences	0.022	0.015	0.018	0.013
Statistics & Probabi- lity	0.011	0.008	0.008	0.004
Substance Abuse	0.000	0.000	0.000	0.000
Surgery	0.784	0.466	0.432	0.446
Telecommunications	0.006	0.006	0.007	0.005
Theater	0.000	0.000	0.000	0.000
Thermodynamics	0.001	0.001	0.001	0.001
Toxicology	0.027	0.025	0.031	0.023
Transplantation	0.000	0.000	0.000	0.000
Transportation	0.000	0.000	0.000	0.000
Transportation Scien- ce & Technology	0.000	0.000	0.000	0.000
Tropical Medicine	0.001	0.001	0.001	0.001
Urban Studies	0.000	0.000	0.000	0.000
Urology & Nephrology	0.001	0.001	0.001	0.000
Veterinary Sciences	0.017	0.011	0.013	0.011
Virology	0.000	0.000	0.000	0.000

DISCIPLINA	Argentina	Brasil	Chile	Colombia
Water Resources	0.003	0.003	0.004	0.004
Women's Studies	0.000	0.000	0.000	0.000
Zoology	0.106	0.088	0.110	0.061
ICD	85.376	74.982	86.788	86.149

Tabla D.1: Tabla de ponderación de importancia disciplinar para Argentina, Brasil, Chile y Colombia.

DISCIPLINA	Cuba	España	Finlandia	Grecia
Acoustics	0.002	0.002	0.002	0.002
Agricultural Econo- mics & Policy	0.000	0.000	0.000	0.000
Agricultural Enginee- ring	0.001	0.000	0.000	0.001
Agriculture, Dairy & Animal Science	0.000	0.001	0.001	0.001
Agriculture, Multidis- ciplinary	0.006	0.008	0.011	0.008
Agronomy	0.007	0.029	0.037	0.024
Allergy	0.000	0.000	0.000	0.000
Anatomy & Morpho- logy	0.000	0.000	0.001	0.000
Andrology	0.000	0.000	0.000	0.000
Anesthesiology	0.000	0.000	0.000	0.000
Anthropology	0.002	0.002	0.003	0.003
Archaeology	0.000	0.000	0.000	0.000
Architecture	0.000	0.000	0.000	0.000
Area Studies	0.000	0.000	0.000	0.000
Art	0.000	0.000	0.000	0.000
Asian Studies	0.000	0.000	0.000	0.000
Astronomy & As- trophysics	0.218	0.183	0.192	0.179
Audiology & Speech- Language Pathology	0.000	0.000	0.000	0.000
Automation & Control Systems	0.002	0.005	0.004	0.005
Behavioral Sciences	0.001	0.001	0.001	0.001
Biochemical Research Methods	0.025	0.046	0.040	0.039
Biochemistry & Mole- cular Biology	23.968	31.552	37.879	27.565
Biodiversity Conserva- tion	0.001	0.002	0.002	0.001

DISCIPLINA	Cuba	España	Finlandia	Grecia
Biology	0.034	0.091	0.096	0.067
Biophysics	0.007	0.009	0.009	0.007
Biotechnology & Applied Microbiology	1.630	2.926	3.112	2.562
Business	0.001	0.002	0.003	0.003
Business, Finance	0.000	0.000	0.000	0.000
Cardiac & Cardiovascular Systems	0.006	0.028	0.036	0.026
Cell & Tissue Engineering	0.000	0.000	0.000	0.000
Cell Biology	0.488	0.918	1.055	0.762
Chemistry, Analytical	0.181	0.354	0.380	0.323
Chemistry, Applied	0.070	0.096	0.097	0.108
Chemistry, Inorganic & Nuclear	0.007	0.013	0.010	0.013
Chemistry, Medicinal	0.009	0.012	0.011	0.012
Chemistry, Multidisciplinary	1.428	3.095	2.971	2.579
Chemistry, Organic	0.001	0.001	0.001	0.001
Chemistry, Physical	8.536	15.929	14.314	16.377
Classics	0.000	0.000	0.000	0.000
Clinical Neurology	0.117	0.180	0.247	0.162
Communication	0.000	0.001	0.001	0.001
Computer Science, Artificial Intelligence	0.039	0.065	0.075	0.065
Computer Science, Cybernetics	0.000	0.000	0.000	0.000
Computer Science, Hardware & Architecture	0.000	0.000	0.000	0.000
Computer Science, Information Systems	0.005	0.011	0.011	0.012
Computer Science, Interdisciplinary Applications	0.636	1.227	1.295	1.097
Computer Science, Software Engineering	0.000	0.000	0.001	0.001
Computer Science, Theory & Methods	0.002	0.004	0.004	0.004
Construction & Building Technology	0.002	0.001	0.002	0.002
Criminology & Penology	0.000	0.000	0.000	0.000

DISCIPLINA	Cuba	España	Finlandia	Grecia
Critical Care Medicine	0.000	0.000	0.000	0.000
Crystallography	0.000	0.000	0.000	0.000
Cultural Studies	0.000	0.000	0.000	0.000
Demography	0.000	0.000	0.000	0.000
Dentistry, Oral Surgery & Medicine	0.012	0.013	0.016	0.013
Dermatology	0.001	0.001	0.001	0.001
Developmental Biology	0.000	0.000	0.000	0.000
Ecology	1.463	2.240	2.386	1.985
Economics	0.226	0.389	0.430	0.471
Education & Educational Research	0.001	0.043	0.097	0.107
Education, Scientific Disciplines	0.000	0.000	0.000	0.000
Education, Special	0.000	0.000	0.000	0.000
Electrochemistry	0.001	0.001	0.002	0.001
Emergency Medicine	0.000	0.000	0.000	0.000
Endocrinology & Metabolism	0.008	0.016	0.019	0.014
Energy & Fuels	1.230	1.301	1.111	1.529
Engineering, Aerospace	0.000	0.000	0.000	0.000
Engineering, Biomedical	0.030	0.042	0.046	0.035
Engineering, Chemical	0.907	1.297	0.922	1.313
Engineering, Civil	0.047	0.144	0.152	0.162
Engineering, Electrical & Electronic	1.685	4.586	4.380	4.736
Engineering, Environmental	0.003	0.003	0.003	0.004
Engineering, Geological	0.000	0.000	0.000	0.000
Engineering, Industrial	0.000	0.000	0.000	0.000
Engineering, Manufacturing	0.000	0.000	0.000	0.000
Engineering, Marine	0.000	0.000	0.000	0.000
Engineering, Mechanical	0.056	0.029	0.029	0.030
Engineering, Multidisciplinary	0.001	0.003	0.004	0.004
Engineering, Ocean	0.000	0.000	0.000	0.000

DISCIPLINA	Cuba	España	Finlandia	Grecia
Engineering, Petro- leum	0.000	0.000	0.000	0.000
Entomology	0.000	0.001	0.001	0.001
Environmental Scien- ces	11.645	20.770	19.541	17.195
Environmental Studies	0.054	0.059	0.060	0.062
Ergonomics	0.000	0.000	0.000	0.000
Ethics	0.000	0.000	0.000	0.000
Ethnic Studies	0.000	0.000	0.000	0.000
Evolutionary Biology	0.000	0.000	0.000	0.000
Family Studies	0.000	0.000	0.000	0.000
Film, Radio, Televi- sion	0.000	0.000	0.000	0.000
Fisheries	0.003	0.005	0.005	0.005
Folklore	0.000	0.000	0.000	0.000
Food Science & Tech- nology	0.340	0.528	0.529	0.499
Forestry	0.001	0.007	0.007	0.005
Gastroenterology & Hepatology	0.001	0.001	0.001	0.001
Genetics & Heredity	0.526	1.121	1.734	1.488
Geochemistry & Geophysics	0.002	0.004	0.003	0.002
Geography	0.000	0.000	0.000	0.000
Geography, Physical	0.009	0.014	0.016	0.014
Geology	0.000	0.000	0.000	0.000
Geosciences, Multidis- ciplinary	0.280	0.702	0.794	0.641
Geriatrics & Geronto- logy	0.002	0.003	0.003	0.002
Gerontology	0.000	0.000	0.000	0.000
GREEN & SUSTAI- NABLE SCIENCE & TECHNOLOGY	0.000	0.000	0.000	0.000
Health Care Sciences & Services	0.002	0.002	0.001	0.002
Health Policy & Servi- ces	0.000	0.000	0.000	0.000
Hematology	0.001	0.001	0.001	0.001
History	0.000	0.001	0.001	0.001
History & Philosophy Of Science	0.000	0.000	0.000	0.000

DISCIPLINA	Cuba	España	Finlandia	Grecia
History Of Social Sciences	0.000	0.000	0.000	0.000
Horticulture	0.000	0.000	0.000	0.000
Hospitality, Leisure, Sport & Tourism	0.000	0.000	0.001	0.001
Humanities, Multidisciplinary	0.000	0.000	0.000	0.000
Imaging Science & Photographic Technology	0.000	0.001	0.001	0.000
Immunology	0.957	1.283	1.578	1.350
Industrial Relations & Labor	0.000	0.000	0.000	0.000
Infectious Diseases	0.091	0.085	0.098	0.102
Information Science & Library Science	0.000	0.000	0.000	0.000
Instruments & Instrumentation	0.025	0.034	0.033	0.030
Integrative & Complementary Medicine	0.000	0.000	0.000	0.000
International Relations	0.000	0.000	0.000	0.000
Language & Linguistics	0.000	0.000	0.000	0.000
Law	0.000	0.000	0.000	0.000
Limnology	0.000	0.000	0.000	0.000
Linguistics	0.000	0.001	0.001	0.001
Literary Reviews	0.000	0.000	0.000	0.000
Literary Theory & Criticism	0.000	0.000	0.000	0.000
Literature	0.000	0.000	0.000	0.000
Literature, African, Australian, Canadian	0.000	0.000	0.000	0.000
Literature, American	0.000	0.000	0.000	0.000
Literature, British Isles	0.000	0.000	0.000	0.000
Literature, German, Dutch, Scandinavian	0.000	0.000	0.000	0.000
Literature, Romance	0.000	0.000	0.000	0.000
Literature, Slavic	0.000	0.000	0.000	0.000
Logic	0.000	0.000	0.000	0.000
Management	0.006	0.021	0.025	0.019

DISCIPLINA	Cuba	España	Finlandia	Grecia
Marine & Freshwater Biology	0.024	0.051	0.053	0.055
Materials Science, Bio-materials	0.000	0.001	0.001	0.001
Materials Science, Ceramics	0.000	0.000	0.000	0.000
Materials Science, Characterization & Testing	0.000	0.000	0.000	0.000
Materials Science, Coatings & Films	0.000	0.000	0.000	0.000
Materials Science, Composites	0.000	0.000	0.000	0.000
Materials Science, Multidisciplinary	8.316	12.752	12.028	11.454
Materials Science, Paper & Wood	0.000	0.000	0.000	0.000
Materials Science, Textiles	0.000	0.000	0.000	0.000
Mathematical & Computational Biology	0.002	0.003	0.003	0.002
Mathematics	0.000	0.000	0.000	0.000
Mathematics, Applied	0.028	0.051	0.049	0.056
Mathematics, Interdisciplinary Applications	0.005	0.011	0.012	0.012
Mechanics	0.033	0.044	0.037	0.053
Medical Ethics	0.000	0.000	0.000	0.000
Medical Informatics	0.000	0.000	0.000	0.000
Medical Laboratory Technology	0.000	0.000	0.000	0.000
Medicine, General & Internal	0.091	0.655	1.281	0.675
Medicine, Legal	0.000	0.000	0.000	0.000
Medicine, Research & Experimental	0.044	0.073	0.080	0.050
Medieval & Renaissance Studies	0.000	0.000	0.000	0.000
Metallurgy & Metallurgical Engineering	0.000	0.000	0.000	0.000
Meteorology & Atmospheric Sciences	0.009	0.017	0.020	0.018
Microbiology	0.106	0.123	0.141	0.148

DISCIPLINA	Cuba	España	Finlandia	Grecia
Microscopy	0.000	0.000	0.000	0.000
Mineralogy	0.000	0.000	0.000	0.000
Mining & Mineral Processing	0.000	0.000	0.000	0.000
Multidisciplinary Sciences	0.228	0.793	1.093	0.790
Music	0.000	0.000	0.000	0.000
Mycology	0.000	0.001	0.001	0.000
Nanoscience & Nanotechnology	0.057	0.108	0.103	0.093
Neuroimaging	0.000	0.000	0.000	0.000
Neurosciences	4.336	4.740	5.638	3.732
Nuclear Science & Technology	0.008	0.012	0.013	0.013
Nursing	0.000	0.001	0.002	0.001
Nutrition & Dietetics	0.066	0.093	0.127	0.119
Obstetrics & Gynecology	0.016	0.025	0.027	0.022
Oceanography	0.003	0.010	0.009	0.009
Oncology	1.352	1.525	1.320	1.056
Operations Research & Management Science	0.016	0.024	0.025	0.027
Ophthalmology	0.000	0.000	0.000	0.000
Optics	0.010	0.030	0.029	0.032
Ornithology	0.000	0.000	0.000	0.000
Orthopedics	0.000	0.000	0.000	0.000
Otorhinolaryngology	0.000	0.000	0.000	0.000
Paleontology	0.000	0.000	0.000	0.000
Parasitology	0.000	0.001	0.001	0.001
Pathology	0.001	0.004	0.005	0.003
Pediatrics	0.044	0.074	0.123	0.077
Peripheral Vascular Disease	0.003	0.003	0.004	0.003
Pharmacology & Pharmacy	1.384	2.369	2.961	2.548
Philosophy	0.000	0.000	0.000	0.000
Physics, Applied	0.052	0.121	0.098	0.110
Physics, Atomic, Molecular & Chemical	0.009	0.020	0.018	0.016
Physics, Condensed Matter	0.004	0.013	0.011	0.011
Physics, Fluids & Plasmas	0.000	0.000	0.000	0.000

DISCIPLINA	Cuba	España	Finlandia	Grecia
Physics, Mathematical	0.004	0.007	0.008	0.007
Physics, Multidisciplinary	0.115	0.160	0.197	0.157
Physics, Nuclear	0.013	0.007	0.009	0.012
Physics, Particles & Fields	0.064	0.044	0.050	0.052
Physiology	0.006	0.009	0.009	0.007
Planning & Development	0.000	0.000	0.000	0.000
Plant Sciences	0.674	1.263	1.442	0.906
Poetry	0.000	0.000	0.000	0.000
Political Science	0.000	0.001	0.001	0.001
Polymer Science	0.037	0.060	0.079	0.065
Primary Health Care	0.000	0.000	0.000	0.000
Psychiatry	0.405	0.563	0.600	0.447
Psychology	0.003	0.005	0.007	0.005
Psychology, Applied	0.000	0.000	0.001	0.001
Psychology, Biological	0.000	0.000	0.000	0.000
Psychology, Clinical	0.000	0.000	0.000	0.000
Psychology, Developmental	0.001	0.001	0.002	0.001
Psychology, Educational	0.000	0.000	0.000	0.000
Psychology, Experimental	0.005	0.006	0.007	0.005
Psychology, Mathematical	0.000	0.000	0.000	0.000
Psychology, Multidisciplinary	0.001	0.002	0.004	0.002
Psychology, Psychoanalysis	0.000	0.000	0.000	0.000
Psychology, Social	0.000	0.000	0.001	0.001
Public Administration	0.000	0.000	0.000	0.000
Public, Environmental & Occupational Health	2.925	2.742	3.125	3.888
Radiology, Nuclear Medicine & Medical Imaging	0.203	0.127	0.164	0.125
Rehabilitation	0.001	0.003	0.004	0.003
Religion	0.000	0.000	0.000	0.000
Remote Sensing	0.000	0.000	0.000	0.000
Reproductive Biology	0.000	0.001	0.001	0.001
Respiratory System	0.001	0.001	0.001	0.001

DISCIPLINA	Cuba	España	Finlandia	Grecia
Rheumatology	0.000	0.000	0.000	0.000
Robotics	0.000	0.000	0.000	0.000
Social Issues	0.000	0.000	0.000	0.000
Social Sciences, Bio-medical	0.000	0.000	0.000	0.000
Social Sciences, Inter-disciplinary	0.000	0.002	0.003	0.002
Social Sciences, Mathematical Methods	0.000	0.000	0.000	0.000
Social Work	0.000	0.000	0.000	0.000
Sociology	0.002	0.001	0.002	0.002
Soil Science	0.000	0.000	0.000	0.000
Spectroscopy	0.000	0.000	0.000	0.000
Sport Sciences	0.012	0.023	0.038	0.028
Statistics & Probability	0.009	0.014	0.012	0.009
Substance Abuse	0.000	0.000	0.000	0.000
Surgery	0.503	0.627	0.820	0.589
Telecommunications	0.003	0.009	0.010	0.010
Theater	0.000	0.000	0.000	0.000
Thermodynamics	0.001	0.001	0.001	0.002
Toxicology	0.020	0.037	0.035	0.039
Transplantation	0.000	0.000	0.000	0.000
Transportation	0.000	0.000	0.000	0.000
Transportation Science & Technology	0.000	0.000	0.000	0.000
Tropical Medicine	0.001	0.001	0.001	0.001
Urban Studies	0.000	0.000	0.000	0.000
Urology & Nephrology	0.000	0.001	0.002	0.001
Veterinary Sciences	0.014	0.022	0.022	0.015
Virology	0.000	0.000	0.000	0.000
Water Resources	0.003	0.006	0.005	0.005
Women's Studies	0.000	0.000	0.000	0.000
Zoology	0.067	0.154	0.188	0.140
ICD	78.396	121.199	128.100	111.565

Tabla D.2: Tabla de ponderación de importancia disciplinar para Cuba, España, Finlandia y Grecia.

DISCIPLINA	Islandia	Israel	Nueva Zelandia	Polonia
Acoustics	0.003	0.002	0.001	0.001

DISCIPLINA		Islandia	Israel	Nueva Zelandia	Polonia
Agricultural Economics & Policy	Econo-	0.000	0.000	0.000	0.000
Agricultural Engineering	Enginee-	0.001	0.000	0.000	0.000
Agriculture, Dairy & Animal Science		0.001	0.001	0.001	0.000
Agriculture, Multidis-		0.005	0.010	0.007	0.006
ciplinary					
Agronomy		0.012	0.035	0.026	0.015
Allergy		0.000	0.000	0.000	0.000
Anatomy & Morpho-		0.001	0.001	0.001	0.000
logy					
Andrology		0.000	0.000	0.000	0.000
Anesthesiology		0.000	0.000	0.000	0.000
Anthropology		0.002	0.003	0.002	0.001
Archaeology		0.000	0.000	0.000	0.000
Architecture		0.000	0.000	0.000	0.000
Area Studies		0.000	0.000	0.000	0.000
Art		0.000	0.000	0.000	0.000
Asian Studies		0.000	0.000	0.000	0.000
Astronomy & As-		0.314	0.192	0.181	0.183
trophysics					
Audiology & Speech-		0.000	0.000	0.000	0.000
Language Pathology					
Automation & Control		0.005	0.005	0.006	0.003
Systems					
Behavioral Sciences		0.001	0.001	0.001	0.001
Biochemical Research		0.052	0.042	0.054	0.028
Methods					
Biochemistry & Mole-		47.305	40.269	34.602	17.751
cular Biology					
Biodiversity Conserva-		0.001	0.001	0.002	0.001
tion					
Biology		0.057	0.093	0.116	0.034
Biophysics		0.010	0.011	0.008	0.006
Biotechnology & Ap-		5.380	3.628	3.138	1.512
plied Microbiology					
Business		0.003	0.003	0.003	0.001
Business, Finance		0.000	0.000	0.000	0.000
Cardiac & Cardiovas-		0.063	0.028	0.032	0.020
cular Systems					

DISCIPLINA	Islandia	Israel	Nueva Zelandia	Polonia
Cell & Tissue Engineering	0.000	0.000	0.000	0.000
Cell Biology	0.916	1.092	0.758	0.406
Chemistry, Analytical	0.650	0.322	0.293	0.222
Chemistry, Applied	0.121	0.091	0.081	0.058
Chemistry, Inorganic & Nuclear	0.007	0.014	0.012	0.007
Chemistry, Medicinal	0.011	0.012	0.013	0.008
Chemistry, Multidisciplinary	2.569	3.734	2.575	1.021
Chemistry, Organic	0.001	0.001	0.001	0.001
Chemistry, Physical	13.173	16.867	15.375	8.825
Classics	0.000	0.000	0.000	0.000
Clinical Neurology	0.256	0.184	0.181	0.128
Communication	0.001	0.002	0.001	0.001
Computer Science, Artificial Intelligence	0.161	0.089	0.068	0.036
Computer Science, Cybernetics	0.000	0.000	0.000	0.000
Computer Science, Hardware & Architecture	0.000	0.000	0.000	0.000
Computer Science, Information Systems	0.014	0.013	0.008	0.008
Computer Science, Interdisciplinary Applications	0.697	1.077	1.817	0.795
Computer Science, Software Engineering	0.001	0.001	0.000	0.000
Computer Science, Theory & Methods	0.002	0.003	0.003	0.003
Construction & Building Technology	0.002	0.002	0.001	0.001
Criminology & Penology	0.000	0.000	0.000	0.000
Critical Care Medicine	0.001	0.000	0.000	0.000
Crystallography	0.000	0.000	0.000	0.000
Cultural Studies	0.000	0.000	0.000	0.000
Demography	0.000	0.000	0.000	0.000
Dentistry, Oral Surgery & Medicine	0.059	0.012	0.014	0.009
Dermatology	0.003	0.001	0.001	0.001

DISCIPLINA		Islandia	Israel	Nueva Zelandia	Polonia
Developmental Biology	Bio-	0.000	0.000	0.000	0.000
Ecology		1.627	1.931	2.504	1.465
Economics		0.374	0.450	0.533	0.165
Education & Educational Research		0.064	0.091	0.086	0.031
Education, Scientific Disciplines	Scientific	0.000	0.000	0.000	0.000
Education, Special		0.000	0.000	0.000	0.000
Electrochemistry		0.002	0.002	0.001	0.001
Emergency Medicine		0.000	0.000	0.000	0.000
Endocrinology & Metabolism		0.015	0.013	0.019	0.009
Energy & Fuels		0.669	1.253	1.122	0.786
Engineering, Aerospace	Aerospace	0.000	0.000	0.000	0.000
Engineering, Biomedical	Biomedical	0.103	0.051	0.036	0.020
Engineering, Chemical		0.557	1.045	0.981	0.439
Engineering, Civil		0.109	0.169	0.129	0.083
Engineering, Electrical & Electronic		9.003	5.458	4.443	1.851
Engineering, Environmental	Environmental	0.003	0.003	0.003	0.001
Engineering, Geological	Geological	0.000	0.000	0.000	0.000
Engineering, Industrial	Industrial	0.000	0.000	0.000	0.000
Engineering, Manufacturing	Manufacturing	0.001	0.000	0.001	0.000
Engineering, Marine		0.000	0.000	0.000	0.000
Engineering, Mechanical		0.018	0.027	0.044	0.020
Engineering, Multidisciplinary		0.002	0.004	0.006	0.003
Engineering, Ocean		0.000	0.000	0.000	0.000
Engineering, Petroleum	Petroleum	0.000	0.000	0.000	0.000
Entomology		0.000	0.001	0.001	0.000
Environmental Sciences	Sciences	21.755	17.358	20.482	9.085
Environmental Studies		0.049	0.049	0.065	0.027

DISCIPLINA	Islandia	Israel	Nueva Zelandia	Polonia
Ergonomics	0.000	0.000	0.000	0.000
Ethics	0.000	0.000	0.000	0.000
Ethnic Studies	0.000	0.000	0.000	0.000
Evolutionary Biology	0.000	0.000	0.001	0.000
Family Studies	0.000	0.000	0.000	0.000
Film, Radio, Television	0.000	0.000	0.000	0.000
Fisheries	0.004	0.005	0.004	0.003
Folklore	0.000	0.000	0.000	0.000
Food Science & Technology	0.513	0.522	0.464	0.294
Forestry	0.008	0.007	0.006	0.002
Gastroenterology & Hepatology	0.001	0.001	0.001	0.001
Genetics & Heredity	4.744	1.112	1.441	0.802
Geochemistry & Geophysics	0.006	0.003	0.004	0.002
Geography	0.000	0.000	0.000	0.000
Geography, Physical	0.015	0.014	0.015	0.009
Geology	0.000	0.000	0.000	0.000
Geosciences, Multidisciplinary	0.669	0.701	0.733	0.433
Geriatrics & Gerontology	0.003	0.002	0.002	0.002
Gerontology	0.000	0.000	0.000	0.000
GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY	0.000	0.000	0.000	0.000
Health Care Sciences & Services	0.001	0.002	0.002	0.001
Health Policy & Services	0.000	0.000	0.000	0.000
Hematology	0.001	0.001	0.001	0.001
History	0.003	0.001	0.002	0.000
History & Philosophy Of Science	0.000	0.000	0.000	0.000
History Of Social Sciences	0.000	0.000	0.000	0.000
Horticulture	0.000	0.000	0.000	0.000
Hospitality, Leisure, Sport & Tourism	0.001	0.001	0.001	0.000

DISCIPLINA	Islandia	Israel	Nueva Zelandia	Polonia
Humanities, Multidis- ciplinary	0.000	0.000	0.000	0.000
Imaging Science & Photographic Techno- logy	0.002	0.001	0.001	0.000
Immunology	1.860	1.686	1.386	0.863
Industrial Relations & Labor	0.000	0.000	0.000	0.000
Infectious Diseases	0.117	0.093	0.093	0.076
Information Science & Library Science	0.000	0.000	0.000	0.000
Instruments & Instru- mentation	0.020	0.025	0.051	0.031
Integrative & Comple- mentary Medicine	0.000	0.000	0.000	0.000
International Rela- tions	0.000	0.000	0.000	0.000
Language & Linguis- tics	0.000	0.000	0.000	0.000
Law	0.000	0.000	0.000	0.000
Limnology	0.000	0.000	0.000	0.000
Linguistics	0.001	0.001	0.001	0.000
Literary Reviews	0.000	0.000	0.000	0.000
Literary Theory & Cri- ticism	0.000	0.000	0.000	0.000
Literature	0.000	0.000	0.000	0.000
Literature, African, Australian, Canadian	0.000	0.000	0.000	0.000
Literature, American	0.000	0.000	0.000	0.000
Literature, British Is- les	0.000	0.000	0.000	0.000
Literature, German, Dutch, Scandinavian	0.000	0.000	0.000	0.000
Literature, Romance	0.000	0.000	0.000	0.000
Literature, Slavic	0.000	0.000	0.000	0.000
Logic	0.000	0.000	0.000	0.000
Management	0.031	0.024	0.020	0.008
Marine & Freshwater Biology	0.045	0.053	0.048	0.031
Materials Science, Bio- materials	0.001	0.001	0.001	0.000

DISCIPLINA	Islandia	Israel	Nueva Zelandia	Polonia
Materials Science, Ce- ramics	0.000	0.000	0.000	0.000
Materials Science, Characterization & Testing	0.000	0.000	0.000	0.000
Materials Science, Coatings & Films	0.000	0.000	0.000	0.000
Materials Science, Composites	0.000	0.000	0.000	0.000
Materials Science, Multidisciplinary	9.915	14.689	10.738	6.625
Materials Science, Pa- per & Wood	0.000	0.000	0.000	0.000
Materials Science, Tex- tiles	0.000	0.000	0.000	0.000
Mathematical & Computational Bio- logy	0.002	0.002	0.004	0.002
Mathematics	0.000	0.000	0.000	0.000
Mathematics, Applied	0.106	0.056	0.043	0.040
Mathematics, Interdis- ciplinary Applications	0.008	0.010	0.015	0.008
Mechanics	0.017	0.035	0.062	0.029
Medical Ethics	0.000	0.000	0.000	0.000
Medical Informatics	0.000	0.000	0.000	0.000
Medical Laboratory Technology	0.000	0.000	0.000	0.000
Medicine, General & Internal	0.876	0.637	0.937	0.740
Medicine, Legal	0.000	0.000	0.000	0.000
Medicine, Research & Experimental	0.078	0.082	0.054	0.025
Medieval & Renaissan- ce Studies	0.000	0.000	0.000	0.000
Metallurgy & Meta- llurgical Engineering	0.000	0.000	0.000	0.000
Meteorology & At- mospheric Sciences	0.019	0.017	0.018	0.012
Microbiology	0.138	0.149	0.133	0.060
Microscopy	0.000	0.000	0.000	0.000
Mineralogy	0.000	0.000	0.000	0.000

DISCIPLINA	Islandia	Israel	Nueva Zelandia	Polonia
Mining & Mineral Processing	0.000	0.000	0.000	0.000
Multidisciplinary Sciences	2.234	1.030	0.826	0.584
Music	0.000	0.000	0.000	0.000
Mycology	0.000	0.001	0.001	0.000
Nanoscience & Nanotechnology	0.076	0.122	0.077	0.061
Neuroimaging	0.000	0.000	0.000	0.000
Neurosciences	5.862	5.269	4.337	3.108
Nuclear Science & Technology	0.009	0.011	0.020	0.013
Nursing	0.002	0.001	0.001	0.001
Nutrition & Dietetics	0.099	0.122	0.090	0.087
Obstetrics & Gynecology	0.016	0.021	0.024	0.009
Oceanography	0.012	0.008	0.009	0.005
Oncology	1.363	1.319	1.187	1.304
Operations Research & Management Science	0.071	0.020	0.025	0.018
Ophthalmology	0.001	0.000	0.000	0.000
Optics	0.029	0.031	0.033	0.022
Ornithology	0.000	0.000	0.000	0.000
Orthopedics	0.001	0.000	0.000	0.000
Otorhinolaryngology	0.000	0.000	0.000	0.000
Paleontology	0.000	0.000	0.000	0.000
Parasitology	0.000	0.001	0.001	0.000
Pathology	0.002	0.004	0.004	0.002
Pediatrics	0.113	0.088	0.102	0.086
Peripheral Vascular Disease	0.005	0.003	0.004	0.003
Pharmacology & Pharmacy	2.646	2.894	2.512	1.609
Philosophy	0.000	0.000	0.000	0.000
Physics, Applied	0.105	0.129	0.115	0.072
Physics, Atomic, Molecular & Chemical	0.020	0.021	0.020	0.014
Physics, Condensed Matter	0.007	0.012	0.011	0.006
Physics, Fluids & Plasmas	0.000	0.000	0.000	0.000
Physics, Mathematical	0.006	0.007	0.007	0.006

DISCIPLINA	Islandia	Israel	Nueva Zelandia	Polonia
Physics, Multidisciplinary	0.171	0.161	0.167	0.071
Physics, Nuclear	0.004	0.009	0.019	0.007
Physics, Particles & Fields	0.036	0.045	0.061	0.044
Physiology	0.006	0.008	0.009	0.007
Planning & Development	0.000	0.000	0.000	0.000
Plant Sciences	0.856	1.434	1.227	0.725
Poetry	0.000	0.000	0.000	0.000
Political Science	0.000	0.001	0.001	0.000
Polymer Science	0.069	0.063	0.055	0.035
Primary Health Care	0.000	0.000	0.000	0.000
Psychiatry	0.697	0.525	0.622	0.256
Psychology	0.008	0.008	0.008	0.005
Psychology, Applied	0.001	0.001	0.001	0.000
Psychology, Biological	0.000	0.000	0.000	0.000
Psychology, Clinical	0.000	0.000	0.000	0.000
Psychology, Developmental	0.002	0.001	0.001	0.001
Psychology, Educational	0.000	0.000	0.000	0.000
Psychology, Experimental	0.007	0.007	0.006	0.005
Psychology, Mathematical	0.000	0.000	0.000	0.000
Psychology, Multidisciplinary	0.003	0.003	0.003	0.002
Psychology, Psychoanalysis	0.000	0.000	0.000	0.000
Psychology, Social	0.001	0.001	0.001	0.000
Public Administration	0.000	0.000	0.000	0.000
Public, Environmental & Occupational Health	3.305	2.496	2.755	1.685
Radiology, Nuclear Medicine & Medical Imaging	0.132	0.151	0.106	0.087
Rehabilitation	0.004	0.003	0.004	0.002
Religion	0.000	0.000	0.000	0.000
Remote Sensing	0.001	0.000	0.000	0.000
Reproductive Biology	0.001	0.001	0.001	0.000
Respiratory System	0.001	0.001	0.001	0.001

DISCIPLINA	Islandia	Israel	Nueva Zelandia	Polonia
Rheumatology	0.000	0.000	0.000	0.000
Robotics	0.000	0.000	0.000	0.000
Social Issues	0.000	0.000	0.000	0.000
Social Sciences, Bio-medical	0.000	0.000	0.000	0.000
Social Sciences, Inter-disciplinary	0.006	0.005	0.003	0.002
Social Sciences, Mathematical Methods	0.000	0.000	0.000	0.000
Social Work	0.000	0.000	0.000	0.000
Sociology	0.002	0.002	0.003	0.000
Soil Science	0.000	0.000	0.000	0.000
Spectroscopy	0.000	0.000	0.000	0.000
Sport Sciences	0.046	0.026	0.036	0.013
Statistics & Probability	0.008	0.010	0.015	0.007
Substance Abuse	0.000	0.000	0.000	0.000
Surgery	0.759	0.578	0.800	0.376
Telecommunications	0.010	0.007	0.007	0.006
Theater	0.000	0.000	0.000	0.000
Thermodynamics	0.001	0.001	0.002	0.001
Toxicology	0.051	0.030	0.036	0.026
Transplantation	0.000	0.000	0.000	0.000
Transportation	0.001	0.001	0.000	0.000
Transportation Science & Technology	0.000	0.000	0.000	0.000
Tropical Medicine	0.001	0.001	0.001	0.000
Urban Studies	0.000	0.000	0.000	0.000
Urology & Nephrology	0.001	0.001	0.001	0.001
Veterinary Sciences	0.023	0.017	0.019	0.009
Virology	0.000	0.000	0.000	0.000
Water Resources	0.004	0.005	0.006	0.003
Women's Studies	0.000	0.000	0.000	0.000
Zoology	0.361	0.132	0.171	0.109
ICD	144.670	132.537	121.678	65.684

Tabla D.3: Tabla de ponderación de importancia disciplinar para Islandia, Israel, Nueva Zelandia y Polonia.

DISCIPLINA	Portugal	Venezuela
Acoustics	0.001	0.001

DISCIPLINA	Portugal	Venezuela
Agricultural Economics & Policy	0.000	0.000
Agricultural Engineering	0.000	0.000
Agriculture, Dairy & Animal Science	0.001	0.000
Agriculture, Multidisciplinary	0.009	0.003
Agronomy	0.031	0.005
Allergy	0.000	0.000
Anatomy & Morphology	0.000	0.000
Andrology	0.000	0.000
Anesthesiology	0.000	0.000
Anthropology	0.002	0.002
Archaeology	0.000	0.000
Architecture	0.000	0.000
Area Studies	0.000	0.000
Art	0.000	0.000
Asian Studies	0.000	0.000
Astronomy & Astrophysics	0.210	0.135
Audiology & Speech-Language Pathology	0.000	0.000
Automation & Control Systems	0.004	0.002
Behavioral Sciences	0.001	0.001
Biochemical Research Methods	0.036	0.027
Biochemistry & Molecular Biology	26.786	28.342
Biodiversity Conservation	0.002	0.001
Biology	0.074	0.038
Biophysics	0.009	0.005
Biotechnology & Applied Microbiology	2.608	2.244
Business	0.002	0.001
Business, Finance	0.000	0.000
Cardiac & Cardiovascular Systems	0.019	0.023
Cell & Tissue Engineering	0.000	0.000

DISCIPLINA	Portugal	Venezuela
Cell Biology	0.911	0.424
Chemistry, Analytical	0.314	0.227
Chemistry, Applied	0.099	0.087
Chemistry, Inorganic & Nuclear	0.013	0.009
Chemistry, Medicinal	0.012	0.006
Chemistry, Multidisci- plinary	2.185	1.789
Chemistry, Organic	0.001	0.001
Chemistry, Physical	14.331	8.964
Classics	0.000	0.000
Clinical Neurology	0.205	0.175
Communication	0.001	0.000
Computer Science, Ar- tificial Intelligence	0.060	0.030
Computer Science, Cy- bernetics	0.000	0.000
Computer Science, Hardware & Architec- ture	0.000	0.000
Computer Science, In- formation Systems	0.010	0.003
Computer Science, In- terdisciplinary Appli- cations	1.083	0.866
Computer Science, Software Engineering	0.000	0.000
Computer Science, Theory & Methods	0.003	0.003
Construction & Build- ing Technology	0.001	0.001
Criminology & Peno- logy	0.000	0.000
Critical Care Medicine	0.001	0.001
Crystallography	0.000	0.000
Cultural Studies	0.000	0.000
Demography	0.000	0.000
Dentistry, Oral Sur- gery & Medicine	0.009	0.002
Dermatology	0.001	0.000
Developmental Bio- logy	0.000	0.000
Ecology	2.004	1.129

DISCIPLINA	Portugal	Venezuela
Economics	0.384	0.251
Education & Educational Research	0.050	0.021
Education, Scientific Disciplines	0.000	0.000
Education, Special	0.000	0.000
Electrochemistry	0.001	0.001
Emergency Medicine	0.000	0.000
Endocrinology & Metabolism	0.012	0.013
Energy & Fuels	1.210	0.666
Engineering, Aerospace	0.000	0.000
Engineering, Biomedical	0.053	0.017
Engineering, Chemical	1.274	0.753
Engineering, Civil	0.152	0.092
Engineering, Electrical & Electronic	4.881	2.792
Engineering, Environmental	0.003	0.003
Engineering, Geological	0.000	0.000
Engineering, Industrial	0.000	0.000
Engineering, Manufacturing	0.001	0.000
Engineering, Marine	0.000	0.000
Engineering, Mechanical	0.032	0.018
Engineering, Multidisciplinary	0.005	0.001
Engineering, Ocean	0.000	0.000
Engineering, Petroleum	0.000	0.000
Entomology	0.001	0.000
Environmental Sciences	16.593	15.238
Environmental Studies	0.037	0.074
Ergonomics	0.000	0.000
Ethics	0.000	0.000
Ethnic Studies	0.000	0.000
Evolutionary Biology	0.000	0.000

DISCIPLINA	Portugal	Venezuela
Family Studies	0.000	0.000
Film, Radio, Television	0.000	0.000
Fisheries	0.005	0.003
Folklore	0.000	0.000
Food Science & Technology	0.513	0.456
Forestry	0.005	0.003
Gastroenterology & Hepatology	0.001	0.001
Genetics & Heredity	0.804	0.974
Geochemistry & Geophysics	0.004	0.002
Geography	0.000	0.000
Geography, Physical	0.012	0.014
Geology	0.000	0.000
Geosciences, Multidisciplinary	0.556	0.527
Geriatrics & Gerontology	0.002	0.004
Gerontology	0.000	0.000
GREEN & SUSTAINABLE SCIENCE & TECHNOLOGY	0.000	0.000
Health Care Sciences & Services	0.001	0.001
Health Policy & Services	0.000	0.000
Hematology	0.001	0.001
History	0.001	0.000
History & Philosophy Of Science	0.000	0.000
History Of Social Sciences	0.000	0.000
Horticulture	0.000	0.000
Hospitality, Leisure, Sport & Tourism	0.000	0.000
Humanities, Multidisciplinary	0.000	0.000
Imaging Science & Photographic Technology	0.001	0.000
Immunology	1.356	1.806

DISCIPLINA	Portugal	Venezuela
Industrial Relations & Labor	0.000	0.000
Infectious Diseases	0.084	0.055
Information Science & Library Science	0.000	0.000
Instruments & Instrumentation	0.033	0.022
Integrative & Complementary Medicine	0.000	0.000
International Relations	0.000	0.000
Language & Linguistics	0.000	0.000
Law	0.000	0.000
Limnology	0.000	0.000
Linguistics	0.001	0.000
Literary Reviews	0.000	0.000
Literary Theory & Criticism	0.000	0.000
Literature	0.000	0.000
Literature, African, Australian, Canadian	0.000	0.000
Literature, American	0.000	0.000
Literature, British Isles	0.000	0.000
Literature, German, Dutch, Scandinavian	0.000	0.000
Literature, Romance	0.000	0.000
Literature, Slavic	0.000	0.000
Logic	0.000	0.000
Management	0.019	0.002
Marine & Freshwater Biology	0.049	0.027
Materials Science, Biomaterials	0.001	0.000
Materials Science, Ceramics	0.000	0.000
Materials Science, Characterization & Testing	0.000	0.000
Materials Science, Coatings & Films	0.000	0.000

DISCIPLINA	Portugal	Venezuela
Materials Science, Composites	0.000	0.000
Materials Science, Multidisciplinary	10.715	7.219
Materials Science, Paper & Wood	0.000	0.000
Materials Science, Textiles	0.000	0.000
Mathematical & Computational Biology	0.002	0.002
Mathematics	0.000	0.000
Mathematics, Applied	0.045	0.038
Mathematics, Interdisciplinary Applications	0.011	0.006
Mechanics	0.052	0.022
Medical Ethics	0.000	0.000
Medical Informatics	0.000	0.000
Medical Laboratory Technology	0.000	0.000
Medicine, General & Internal	0.299	0.229
Medicine, Legal	0.000	0.000
Medicine, Research & Experimental	0.069	0.016
Medieval & Renaissance Studies	0.000	0.000
Metallurgy & Metallurgical Engineering	0.000	0.000
Meteorology & Atmospheric Sciences	0.018	0.006
Microbiology	0.117	0.136
Microscopy	0.000	0.000
Mineralogy	0.000	0.000
Mining & Mineral Processing	0.000	0.000
Multidisciplinary Sciences	0.803	1.674
Music	0.000	0.000
Mycology	0.001	0.000
Nanoscience & Nanotechnology	0.091	0.078
Neuroimaging	0.000	0.000

DISCIPLINA	Portugal	Venezuela
Neurosciences	4.840	3.299
Nuclear Science & Technology	0.013	0.007
Nursing	0.001	0.000
Nutrition & Dietetics	0.094	0.036
Obstetrics & Gynecology	0.019	0.006
Oceanography	0.010	0.008
Oncology	1.209	1.643
Operations Research & Management Science	0.028	0.024
Ophthalmology	0.000	0.001
Optics	0.028	0.017
Ornithology	0.000	0.000
Orthopedics	0.000	0.000
Otorhinolaryngology	0.000	0.000
Paleontology	0.000	0.000
Parasitology	0.001	0.000
Pathology	0.003	0.002
Pediatrics	0.074	0.055
Peripheral Vascular Disease	0.003	0.001
Pharmacology & Pharmacy	2.557	1.989
Philosophy	0.000	0.000
Physics, Applied	0.101	0.062
Physics, Atomic, Molecular & Chemical	0.016	0.012
Physics, Condensed Matter	0.010	0.007
Physics, Fluids & Plasmas	0.000	0.000
Physics, Mathematical	0.006	0.004
Physics, Multidisciplinary	0.161	0.051
Physics, Nuclear	0.010	0.004
Physics, Particles & Fields	0.048	0.035
Physiology	0.007	0.009
Planning & Development	0.000	0.001
Plant Sciences	0.991	0.666
Poetry	0.000	0.000

DISCIPLINA	Portugal	Venezuela
Political Science	0.001	0.000
Polymer Science	0.064	0.080
Primary Health Care	0.000	0.000
Psychiatry	0.451	0.361
Psychology	0.005	0.001
Psychology, Applied	0.001	0.000
Psychology, Biological	0.000	0.000
Psychology, Clinical	0.000	0.000
Psychology, Develop- mental	0.001	0.001
Psychology, Educatio- nal	0.000	0.000
Psychology, Experi- mental	0.005	0.000
Psychology, Mathema- tical	0.000	0.000
Psychology, Multidis- ciplinary	0.002	0.002
Psychology, Psychoa- nalysis	0.000	0.000
Psychology, Social	0.001	0.000
Public Administration	0.000	0.000
Public, Environmental & Occupational Health	1.925	3.116
Radiology, Nuclear Medicine & Medical Imaging	0.100	0.131
Rehabilitation	0.003	0.003
Religion	0.000	0.000
Remote Sensing	0.000	0.000
Reproductive Biology	0.001	0.001
Respiratory System	0.000	0.001
Rheumatology	0.000	0.000
Robotics	0.000	0.000
Social Issues	0.000	0.000
Social Sciences, Bio- medical	0.000	0.000
Social Sciences, Inter- disciplinary	0.002	0.000
Social Sciences, Mat- hematical Methods	0.000	0.000
Social Work	0.000	0.000
Sociology	0.001	0.001

DISCIPLINA	Portugal	Venezuela
Soil Science	0.000	0.000
Spectroscopy	0.000	0.000
Sport Sciences	0.023	0.007
Statistics & Probability	0.010	0.004
Substance Abuse	0.000	0.000
Surgery	0.519	0.596
Telecommunications	0.007	0.007
Theater	0.000	0.000
Thermodynamics	0.001	0.000
Toxicology	0.035	0.023
Transplantation	0.000	0.000
Transportation	0.000	0.000
Transportation Science & Technology	0.000	0.000
Tropical Medicine	0.001	0.001
Urban Studies	0.000	0.000
Urology & Nephrology	0.001	0.002
Veterinary Sciences	0.019	0.006
Virology	0.000	0.000
Water Resources	0.005	0.006
Women's Studies	0.000	0.000
Zoology	0.134	0.094
ICD	104.954	90.198

Tabla D.4: Tabla de ponderación de importancia disciplinar para Portugal y Venezuela.