

Redes Temáticas en la Web

**Estudio de caso de la Red Temática de la Transparencia en
Chile**

**Autor
Miguel Castillejo Sierra**

**TESI DOCTORAL
UNIVERSITAT POMPEU FABRA 2015**

DEPARTAMENT DE COMUNICACIÓ

DIRECTORS

Dr. Lluís Codina

Universitat Pompeu Fabra

Dr. Carles Pont

Universitat Pompeu Fabra



A mi abuelo, mi padre y David Attenborough por enseñarme a amar la naturaleza.

Agradecimientos

Le doy las gracias a Lluís Codina y Carles Pont, mis directores de tesis, por enseñarme con cariño y acogerme como uno más.

A Carlos Gonzalo por ser un compañero y compartir sus ideas conmigo.

A Sabela Fernández por introducirme en el maravilloso mundo de la ciencia.

A Christina Grammatikopoulo y Modesta di Paola por enseñarme a escribir en español y por llevarme a casa en taxi.

A Miquel Comadrán por hacer de Barcelona mi casa.

A mi madre y mi hermana por empujarme siempre.

A Verónica por amarme.

Resumen

El objeto de estudio de esta investigación son las **Redes Temáticas**, concretamente las Redes Temáticas en la Web y su potencial para extraer datos objetivos de las corrientes de opinión que se generan en torno a un tema de discusión o controversia social.

Esta investigación se estructura a través de cuatro objetivos: caracterizar los componentes de las redes temáticas; caracterizar y evaluar las herramientas para el análisis de redes temáticas en la web; diseñar un Sistema de Análisis de Redes Temáticas en la Web; y aplicar el Sistema de Análisis al caso de estudio de la Red Temática de la Transparencia en Chile.

Como conclusiones, presentamos y caracterizamos los componentes de una red temática en la web: redes de hiperenlaces, actores y temas; analizamos los resultados de la evaluación de las herramientas que consideramos más adecuadas para el análisis de redes temáticas en la web: IssueCrawler, SocSciBot, Webometric Analyst y VOSON; construimos un sistema de análisis dividido en tres fases: análisis de redes de hiperenlaces, análisis de actores y análisis de temas; y finalmente discutimos los resultados del análisis de la Red Temática de la Transparencia en Chile y los posibles desarrollos futuros de la investigación.

Abstract

The object of study of this research are Issue Networks, namely the Issue networks that are active within the domain of the Internet and their potential to extract objective data from the opinion flows that are generated in regard to an issue of discussion or social controversy.

This research is founded on four objectives: the characterization of the components of issue networks; the identification, description and evaluation of existing tools for the analysis of issue networks on the Internet; creation of an Analysis System of Issue Networks on the Internet; and, lastly, the application of the Analysis System to the case study of the Issue Network for Transparency in Chile.

In conclusion, we introduce the characteristics of the components of an Issue Network on the Internet: hyperlinks, actors and issue networks; we present the results of the evaluation of the tools that we consider most suitable for the analysis of Issue Networks on the Internet: IssueCrawler, SocSciBot, Webometric Analyst and VOSON; we build an analysis system divided into three parts: network analysis of hyperlinks, stakeholder analysis and issue analysis; and finally we discuss the results of the analysis of the Issue Network for Transparency in Chile and the possible future developments of the investigation.

ÍNDICE

INTRODUCCIÓN

1.1. Objeto de estudio y objetivos	1
1.1.1. Primer objetivo: Definir y caracterizar los componentes de las redes temáticas en la web.	3
1.1.2. Segundo objetivo: Desarrollar una caracterización y un sistema de evaluación de herramientas para el análisis de redes temáticas en la web	4
1.1.3. Tercer objetivo: Diseñar un sistema de análisis de redes temáticas en la web.	5
1.1.4. Cuarto objetivo: Aplicar el sistema de análisis al caso de la red temática de la transparencia en Chile	6
1.2. Motivación y oportunidad de la investigación	7
1.3. Preguntas de Investigación	17
1.3.1. Preguntas genéricas	17
1.3.2. Preguntas específicas	18
1.4. Marco teórico	20
1.4.1. En relación al primer objetivo de investigación	21
1.4.2. En relación al segundo objetivo de investigación	26
1.4.3. En relación al tercer objetivo de investigación	28
1.4.4. En relación al cuarto objetivo de investigación	31
1.5. Marco metodológico	33
1.5.1. En relación al primer objetivo de investigación	35
1.5.2. En relación al segundo objetivo de investigación	38
1.5.3. En relación al tercer objetivo de investigación	41
1.5.4. En relación al cuarto objetivo de investigación	43
2. REDES TEMÁTICAS EN EL ÁMBITO DE LA COMUNICACIÓN SOCIAL	
2.1. Consideraciones preliminares en torno a las redes temáticas	45
2.2. Redes temáticas en la web: una propuesta de caracterización y componentes	69
2.2.1. Redes de hiperenlaces	73
2.2.2. Temas	75

2.2.3. Actores _____	77
2.2.4. Redes temáticas _____	79
3. HERRAMIENTAS PARA EL ANÁLISIS DE REDES TEMÁTICAS EN LA WEB	
3.1. Consideraciones previas entorno a las herramientas para el análisis de redes temáticas en la web _____	82
3.2. Caracterización y evaluación de herramientas para el análisis de redes temáticas en la web _____	97
3.2.1. Perspectivas y modelos de calidad en la evaluación de software _____	98
3.2.2. Selección y caracterización de herramientas _____	115
a) IssueCrawler _____	115
b) SocSciBot _____	123
c) Webometric Analyst _____	131
d) VOSON _____	140
3.2.3 Propuesta de modelo de calidad y evaluación de herramientas _____	155
4. DISEÑO Y APLICACIÓN DE UN SISTEMA DE ANÁLISIS DE REDES TEMÁTICAS EN LA WEB	
4.1. Diseño de un Sistema de Análisis de Redes Temáticas en la Web _____	159
4.1.1. Fases y componentes del análisis de redes de hiperenlaces _____	160
a) Selección de semillas _____	160
b) Análisis de enlaces _____	162
c) Análisis de redes sociales _____	166
d) Visualización de redes de hiperenlaces _____	169
4.1.2. Fases y componentes del análisis de actores _____	172
a) Identificación de actores _____	172
b) Análisis de redes sociales _____	175
c) Visualización de redes de vecindario _____	177
4.1.3. Fases y componentes para el análisis de temas _____	178
a) Análisis de comunidades _____	181
b) Análisis de actores _____	182
c) Visualización de nubes de palabras _____	183
4.2. Estudio de caso: Análisis de la Red Temática de la Transparencia en Chile _____	184
4.2.1. Análisis de la red de hiperenlaces _____	185

a) Selección de semillas _____	185
b) Rastreo de enlaces _____	188
c) Análisis y Visualización _____	190
4.2.2. Análisis de actores _____	209
a) Identificación de actores _____	209
b) Análisis de redes sociales _____	209
c) Visualización de redes de vecindario _____	210
4.2.3. Análisis de temas _____	238

5. DISCUSIÓN DE RESULTADOS, CONCLUSIONES E INVESTIGACIONES FUTURAS

5.1. Discusión de resultados

5.1.1. Eje de la Definición y Caracterización de Componentes de una Red Temática _____	244
a) Caracterización de una red temática _____	244
b) Componentes de una red temática _____	244
c) Proposiciones conceptuales _____	245
5.1.2. Eje de la caracterización y evaluación de herramientas _____	245
5.1.3. Eje del Sistema de Análisis de Redes Temáticas en la Web _____	246
a) Rendimiento y potencialidad del Sistema de Análisis _____	246
b) Capacidad de identificación _____	249
c) Economía de medios _____	249
d) Fases y Componentes de un Sistema de Análisis de Redes Temáticas _____	249
e) Análisis de Redes y Periodismo de Datos _____	250
f) Fases de Aplicación del Sistema de Análisis _____	250
5.1.4. Eje del Estudio de Caso de la Red Temática de la Transparencia en Chile _____	253
a) Aplicación al análisis de fenómenos de opinión _____	253
b) Caracterización de una red de hiperenlaces _____	253
c) Identificación de actores o comunidades _____	254
d) Identificación de posiciones o marcos de referencia en una Red Temática _____	254
e) Identificación de actores relevantes en una red temática _____	254

f) Identificación de relaciones entre los actores de una red temática _____	255
g) Identificación de posturas de los actores en una red temática _	256
5.2. Conclusiones _____	256
5.3. Investigaciones futuras	
5.3.1. Análisis de plataformas sociales _____	258
5.3.2. Análisis de contenido _____	259
5.3.3. Correlación entre clústers de redes y clústers de contenidos _____	259

LISTA DE FIGURAS

Figura 1. Resumen de los cuatro objetivos de investigación de Redes temáticas en la web: estudio de caso de la red temática de la transparencia en Chile. _____	7
Figura 2. Representación de los avances tecnológicos más importantes de la historia de la humanidad. Extraído de Roca (2012) _____	8
Figura 3. Datos sobre la formas de acceso a las noticias ordenados por países y tipos de plataforma. Extraído de Reuters Institute for the Study of Journalism, (2015) _____	11
Figura 4. Resumen del Marco Teórico de la Comunicación Social en el 1º objetivo de Investigación _____	23
Figura 5. Resumen del Marco Teórico de la Sociología en el 1º objetivo de investigación o Describir y caracterizar los componentes de las redes temáticas en la Web _____	26
Figura 6. Resumen del Marco Teórico de la Evaluación de Calidad Informática en el 2º objetivo de investigación o Identificar, Describir y Evaluar las herramientas para el análisis de redes temáticas en la Web. _____	28
Figura 7. Resumen del Marco Teórico de las Ciencias de la Información en el 3º objetivo de investigación o Diseñar un sistema de análisis de redes temáticas en la Web _____	30
Figura 8. Resumen del Marco Teórico de la Comunicación Social en el 3º objetivo de investigación o Diseñar un sistema de análisis de redes temáticas en la Web _____	31
Figura 9. Resumen del Marco Teórico en el 4º objetivo de investigación o Aplicar el sistema de análisis al estudio de caso de la red temática de la transparencia en Chile _____	32
Figura 10: Modelo comparado de esfera pública en la sociedad de masas y en la sociedad red respectivamente. Extraído de López-García (2000: 236-240) _____	57
Figura 11: Red de mundo pequeño. Extraído de Watts y Strogatz (1998: 441). _____	59
Figura 12: Funciones de distribución de conectividad en grandes redes. (A) Red de colaboración de actores N = 212.250 vértices en promedio de conectividad (k) = 28.78. (B) WWW. N = 325.729 (k) = 5.46 (6). (C) Red eléctrica. N = 4941, (k) = 2.67. Extraído de Barabási y Albert (1999: 510). _____	60
Figura 13. Esquema tipo de un crawler. Elaboración propia. _____	90
Figura 14. Representación gráfica de una red aleatoria y una red libre de escala junto a las funciones numéricas de nodos y enlaces (Perpiñá, 2009). _____	91
Figura 15. Representación gráfica de la estructura de la web según el modelo de pajarita (Broder et al, 2000). _____	93

Figura 16. Triángulo de la calidad de McCall (1977). Elaboración propia. _____	103
Figura 17. Modelo de calidad de Bohem, (1978.) Elaboración propia. _____	104
Figura 18. Modelo genérico de calidad de Dromey (1995). Elaboración propia. _____	107
Figura 19. Modelo de calidad interior y exterior ISO/IEC 9126 (2001). Elaboración propia _____	109
Figura 20. Modelo de calidad en uso ISO/IEC 9126 (2001). Elaboración propia. _____	110
Figura 21. Divisiones del modelo ISO/IEC 25000 SQuaRE (2005). Elaboración propia. ____	111
Figura 22. Modelo de calidad del producto software ISO/IEC 25010 SQuaRE (2011). Elaboración propia. _____	112
Figura 23. Modelo de calidad en uso ISO/IEC 25010 SQuaRE (2011). Elaboración propia. _____	113
Figura 24. Módulo harvester. Extraído de IssueCrawler. _____	117
Figura 25. Módulo de selección de rastreo. Extraído de IssueCrawler. _____	117
Figura 26. Módulo de configuración de rastreo. Extraído de IssueCrawler. _____	118
Figura 27. Módulo de configuración avanzada de rastreo. Extraído de IssueCrawler. ____	119
Figura 28. Módulo de gestión de redes. Extraído de IssueCrawler. _____	120
Figura 29. Módulo de archivo. Extraído de IssueCrawler. _____	121
Figura 30. Módulo de visualización e interactividad. Extraído de IssueCrawler. _____	121
Figura 31. Opciones de configuración del módulo de visualización e interactividad. Extraído de IssueCrawler. _____	122
Figura 32. Ejemplo de visualización redes generado con IssueCrawler. _____	123
Figura 33. Módulo de rastreo. Extraído de SocSciBot. _____	125
Figura 34. Opciones de configuración del rastreo. Extraído de SocSciBot. _____	126
Figura 35. Lista de expresiones regulares. Extraído de SocSciBot. _____	127
Figura 36. Módulo de análisis de enlaces. Extraído de SocSciBot. _____	127
Figura 37. Módulo de informes. Extraído de SocSciBot. _____	129
Figura 38. Módulo de análisis de contenido Cyclist. Extraído de SocSciBot. _____	130

Figura 39. Interfaz avanzada. Extraído de Webometric Analyst. _____	133
Figura 40. Interfaz reducida wizard. Extraído de Webometric Analyst. _____	134
Figura 41. Lista de utilidades. Extraído de Webometric Analyst. _____	137
Figura 42. Ejemplo de visualización de redes generado con Webometric Analyst. _____	139
Figura 43. Opciones de configuración de rastreo. Extraído de VOSON. _____	143
Figura 44. Módulo de selección de semillas. Extraído de VOSON. _____	143
Figura 45. Módulo de visualización de bases de datos. Extraído de VOSON. _____	144
Figura 46. Módulo de visualización de resultados. Extraído de VOSON. _____	145
Figura 47. Módulo de preprocesamiento de datos. Extraído de VOSON. _____	146
Figura 48. Módulo de codificación. Extraído de VOSON. _____	146
Figura 49. Ejemplo de visualización de redes en formato <i>mínimum spanning tree</i> generado con VOSON. _____	148
Figura 50. Ejemplo de visualización de redes en formato completo generado con VOSON. _____	149
Figura 51. Ejemplo de visualización de redes de conceptos generado con VOSON. _____	150
Figura 52. Ejemplo de visualización de redes en formato jerárquico generado con VOSON. _____	151
Figura 53. Propuesta de modelo de evaluación de la calidad general en uso de crawlers. Elaboración propia. _____	153
Figura 54. Propuesta de modelo de evaluación de la calidad funcional en crawlers. Elaboración propia _____	153
Figura 55. Ejemplo de nodo con enlaces salientes. Elaboración propia. _____	162
Figura 56. Ejemplo de nodo con enlaces entrantes. Elaboración propia. _____	163
Figura 57. Ejemplo de nodo con enlaces entrantes. Elaboración propia. _____	163
Figura 58. Ejemplo de nodos conectados indirectamente por coenlace saliente. Elaboración propia. _____	164
Figura 59. Ejemplo de nodos conectados indirectamente por coenlace entrante. Elaboración propia. _____	164
Figura 60. Red Temática Agrupada de la Transparencia en Chile. _____	192

Figura 61. Red Temática de la Transparencia en Chile refinada considerando únicamente enlaces recíprocos y coenlaces y dividida en comunidades por colores. _____	193
Figura 62. Representación gráfica y análisis de redes sociales y de temas del Clúster 1. _	197
Figura 63. Representación gráfica y análisis de redes sociales y de temas del Clúster 2. _	199
Figura 64. Representación gráfica y análisis de redes sociales y de temas del Clúster 3. _	201
Figura 65. Representación gráfica y análisis de redes sociales y de temas del Clúster 4. _	203
Figura 66. Representación gráfica y análisis de redes sociales y de temas del Clúster 5. _	205
Figura 67. Representación gráfica y análisis de redes sociales y de temas del Clúster 6. _	207
Figura 68. Representación gráfica de la red de vecindario, análisis de redes sociales y análisis de temas del actor Comisión Defensora Ciudadana y Transparencia. _____	212
Figura 69. Representación gráfica de la red de vecindario, análisis de redes sociales y análisis de temas del actor Educatransparencia. _____	215
Figura 70. Representación gráfica de la red de vecindario, análisis de redes sociales y análisis de temas del actor Poderopedia. _____	217
Figura 71. Representación gráfica de la red de vecindario, análisis de redes sociales y análisis de temas del actor Consejo de Transparencia _____	220
Figura 72. Representación gráfica de la red de vecindario y análisis de redes sociales de Ciudadano Inteligente. _____	223
Figura 73. Representación gráfica de la red de vecindario y análisis de redes sociales de Gobierno Transparente. _____	226
Figura 74. Representación gráfica de la red de vecindario y análisis de redes sociales de Ley Chile. _____	229
Figura 75. Representación gráfica de la red de vecindario y análisis de redes sociales de Ciudad Viva. _____	231
Figura 76. Representación gráfica de la red de vecindario y análisis de redes sociales de Probono. _____	233
Figura 77. Representación gráfica de la red de vecindario y análisis de redes sociales de Chile Transparente. _____	236

LISTA DE TABLAS

Tabla 1. Palabras clave utilizadas en la revisión sistemática para la construcción del Marco Teórico del 1º objetivo de investigación o Describir y caracterizar los componentes de las redes temáticas en la Web _____	37
Tabla 2. Ecuaciones de búsqueda utilizadas en la revisión sistemática para la construcción del Marco Teórico del 1º objetivo de investigación o Describir y caracterizar los componentes de las redes temáticas en la Web _____	37
Tabla 3. Palabras clave utilizadas en la revisión sistemática para la construcción del Marco Teórico del 2º objetivo de investigación o Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la Web _____	39
Tabla 4. Ecuaciones de búsqueda utilizadas en la revisión sistemática para la construcción del Marco Teórico del 1º objetivo de investigación o Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la Web. _____	39
Tabla 5. Palabras clave utilizadas en la revisión sistemática para la construcción del Marco Teórico del 3º objetivo de investigación o Diseñar un sistema de análisis de redes temáticas en la Web _____	41
Tabla 6. Ecuaciones de búsqueda utilizadas en la revisión sistemática para la construcción del Marco Teórico del 3º objetivo de investigación o Diseñar un sistema de análisis de redes temáticas en la Web _____	42
Tabla 7: Resumen comparado de las teorías de análisis de redes sociales, de análisis hipertexto y de análisis de redes temáticas. Adaptado de Lee (2006:71). _____	68
Tabla 8. Lista de técnicas, objetivos y criterios de selección utilizados en los crawlers (Olston y Najork, 2010). _____	95
Tabla 9. Modelo de calidad FURPS, (Eeles, 2005). Elaboración propia _____	105
Tabla 10. Evaluación de la calidad general en uso de IssueCrawler, SocSciBot, Webometric Analyst y VOSON. Elaboración propia. _____	155
Tabla 11. Evaluación de la calidad funcional de IssueCrawler, SocSciBot, Webometric Analyst y VOSON. Elaboración propia. _____	156
Tabla 12. Semillas obtenidas a través de consultas en buscadores. Elaboración propia _____	185

Tabla 13. Semillas obtenidas a través de consultas en buscadores y directorios. Elaboración propia _____	203
Tabla 14. Medidas de análisis de redes sociales de los clústers 1-6. Elaboración propia _	195
Tabla 15. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 1. _____	197
Tabla 16. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 2. _____	199
Tabla 17. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 3. _____	201
Tabla 18. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 4. _____	203
Tabla 19. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 5. _____	205
Tabla 20. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 6. _____	207
Tabla 21. Medidas de redes sociales de los actores 1-10. _____	209
Tabla 22. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Comisión Defensora Ciudadana y Transparencia. _____	212
Tabla 23. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Comisión Defensora Ciudadana y Transparencia. _____	213
Tabla 24. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Educatransparencia _____	215
Tabla 25. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Educatransparencia _____	215
Tabla 26. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Poderopedia _____	217
Tabla 27. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Poderopedia _____	218

Tabla 28. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Consejo de Transparencia_____	220
Tabla 29. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Consejo Transparencia_____	221
Tabla 30. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Ciudadano Inteligente_____	223
Tabla 31. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Ciudadano Inteligente_____	224
Tabla 32. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Gobierno Transparente_____	226
Tabla 33. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente_____	227
Tabla 34. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente_____	229
Tabla 35. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente_____	231
Tabla 36. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Probono_____	233
Tabla 37. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente_____	234
Tabla 38. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Probono_____	236
Tabla 39. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente_____	236

INTRODUCCIÓN

1.1. Objeto de estudio y objetivos

El objeto de estudio de esta investigación son las **Redes Temáticas**. En particular, en nuestra investigación nos hemos interesado por las características de las redes temáticas presentes en la Web. Es decir, y siempre desde la perspectiva combinada de las ciencias sociales y las ciencias de la información, los conjuntos de individuos u organizaciones que relacionados a través de una red de hiperenlaces plantean opiniones y utilizan recursos u oportunidades para influir en un tema de discusión.

En investigaciones previas presentamos una propuesta articulada de definición de Redes Temáticas donde se sentaron las bases para el proyecto de trabajo de esta tesis. Esta definición constituye el punto de partida de nuestro trabajo y es el resultado de un análisis sistemático de la producción científica sobre este campo que, partiendo de la disciplina general de la Comunicación Social y la Sociología, se centra en el marco teórico del Análisis de Redes Sociales y concretamente del **Análisis de Redes Temáticas en la Web**.

En el modelo conceptual de Redes Temáticas que proponemos, y que la investigación que presentamos aquí ha tenido la oportunidad de validar, sugeríamos la siguiente composición de nuestro objeto de estudio:

- **Redes de hiperenlaces.** En ciencias sociales y en ciencias de la información, una red de hiperenlaces es el sistema dinámico que hace posible el flujo de información entre diferentes sitios web y a través del cual se establecen relaciones de comunidad entre los actores que los crean¹.

¹ Recordemos que entendemos por hiperenlace la porción del código fuente de una página web (elemento <a> más atributo href) que activa un nuevo destino de navegación si el usuario hace clic sobre el mismo.

- **Temas.** Un tema es un conjunto de ideas, teorías o conceptos que generan controversia entre grupos de actores políticos, sociales o económicos².
- **Actores.** Un actor es todo individuo u organización, del ámbito público o privado, que tenga quejas, recursos u oportunidad de influir en un tema de discusión³.

Como tendremos oportunidad de ver en el Capítulo 1, en nuestro modelo conceptual entendemos que una red temática es un conjunto de individuos u organizaciones que plantean opiniones y utilizan recursos u oportunidades para influir en un tema de discusión, y esto siempre de una forma mediada por la web.

La idea esencial para nuestra investigación es la conjetura o hipótesis inicial de que el análisis de redes temáticas permite aportar información sobre las características formales de los temas sociales de discusión y debate en la Web. Por la misma razón, nuestra hipótesis inicial incluye la idea de que el análisis de las redes permite extraer datos objetivos de las corrientes de opinión que se generan en torno a un tema a lo largo del tiempo.

Para acompañar y reforzar los desarrollos principales de esta investigación, las principales Ideas Clave (IC, a partir de ahora) se resumirán al final de cada subapartado y se aportarán las tablas y figuras que se estimen necesarias para el mismo efecto.

² No existe en castellano la diferenciación que existe en inglés entre *topic* e *issue*. En inglés se considera que un *topic* es simplemente una forma de generalización simbólica que posibilita la comunicación, mientras que un *issue* se forma a través de la defensa de posturas diferenciadas en una controversia generada en torno a un tema que podríamos definir como político. Por tanto, los temas políticos o *issues* son objetos de discusión que se desarrollan a través de la interacción entre actores sociales y políticos con distintas posiciones.

³ El concepto de actor que proponemos surge de la literatura de los movimientos sociales (Jenkins, 1995), los grupos de interés (Berry, 1984) y de la literatura de los *stakeholders* o partes interesadas (Freeman, 1984). Consideramos que en los tres ámbitos el fenómeno que se estudia presenta un gran nivel de similitud y es de alguna manera intercambiable: aquellos que defienden una determinada demanda. Utilizamos el término actor en un intento de abarcar estos tres ámbitos utilizando un único término que pueda incluir (1) tanto a los individuos u organizaciones que forman parte de un movimiento social, como (2) un grupo de interés o (3) partes interesadas que actúan como *stakeholders* frente a una empresa privada.

IC1

El principal objetivo del análisis de redes temáticas es extraer datos objetivos de la forma y, eventualmente, de las corrientes de opinión que se generan en torno a un tema a lo largo del tiempo.

1.1.1. Primer objetivo: Definir y caracterizar los componentes de las redes temáticas en la web.

La naturaleza dinámica de la web y la falta de controles de publicación hace muy difícil analizar y extraer conclusiones claras de los datos cuantitativos que las diferentes técnicas de recuperación de información nos permiten extraer de la Web. La dificultad de extraer muestras de información suficientemente grandes y la volatilidad de las muestras extraídas provocada por la constante publicación de información nueva hacen muy difícil desarrollar un sistema de análisis fácilmente operable y ágil que permita llevar a cabo análisis periódicos con un bajo coste de recursos temporales, humanos o económicos.

Para conseguir que el análisis de redes temáticas en la web pueda ser aplicado con una agilidad suficiente y que así se adapte a la dinámica característica de los fenómenos de opinión en la web, es necesario establecer metodologías claras y sistemáticas, que puedan ser: (a) escalables y (b) reproducibles con una periodicidad de tiempo razonablemente corta.

Ahora bien, consideramos que únicamente podremos proponer un sistema de análisis de redes temáticas tras definir claramente cuáles son los elementos que las caracterizan. Por este motivo, el primer objetivo de esta investigación es identificar, describir y, en su caso, validar las características que deben cumplir los elementos que forman parte o intervienen en una red temática y que, de acuerdo con los resultados de nuestras investigaciones previas y como señalábamos anteriormente, son los siguientes: (1) las redes de hiperenlaces que generan la red y dan sentido al término, (2) los temas de discusión que constituyen el “tema” objeto de debate y (3) los actores

que participan en el debate utilizando recursos de la web para hacer públicas sus opiniones.

IC2

El primer objetivo de investigación o **Definir y caracterizar los componentes de las redes temáticas en la web** se responde en el **Capítulo 2** de esta tesis.

IC3

Los elementos que forman parte o intervienen en una red temática son: **(1) las redes de hiperenlaces** que generan la red, **(2) los temas de discusión** que constituyen el “tema” objeto de debate y **(3) los actores** que participan en el debate.

1.1.2. Segundo objetivo: Identificar, Describir y Evaluar las herramientas para el análisis de redes temáticas en la Web.

Con el fin de asegurar la máxima eficacia en el análisis de las Redes Temáticas, nos proponemos el desarrollo de un sistema que nos permita identificar, caracterizar y evaluar las principales herramientas disponibles para el análisis de redes temáticas en la web.

Se trata de un objetivo que corresponde a una fase previa e instrumental al objeto de estudio pero que resulta imprescindible para alcanzar los siguientes objetivos de la investigación. No obstante, entendemos que, pese a este carácter instrumental que señalamos, este sistema de evaluación puede constituir en sí mismo, una aportación de nuestra investigación que puede ser útil para otros investigadores.

En concreto, las herramientas de software necesarias para desarrollar análisis de redes temáticas en la web son los denominados crawlers. En nuestra investigación, los crawlers elegidos para ser caracterizados y evaluados son IssueCrawler, SocSciBot, Webometric Analyst y VOSON.

IC4

El segundo objetivo de investigación o **Identificar, Describir y Evaluar las herramientas para el análisis de redes temáticas en la Web** se responde en el **Capítulo 3** de esta tesis.

IC5

En nuestra investigación, los crawlers elegidos para ser caracterizados y evaluados son **IssueCrawler, SocSciBot, Webometric Analyst** y **VOSON**.

1.1.3. Tercer objetivo: Diseñar un sistema de análisis de redes temáticas en la web.

A través de un estudio sistemático de la producción científica relacionada con el análisis de redes de hiperenlaces y el análisis de redes temáticas se establecerán los criterios que se pueden considerar más adecuados para analizar redes temáticas en la web. A partir de los criterios propuestos se articulará una metodología de análisis que permita aplicar el sistema al caso de estudio de la red temática de la transparencia en Chile.

Consideramos que el análisis de redes temáticas, para ser completo, debe desarrollarse en tres fases que corresponden al análisis de los tres componentes que forman parte de una red temática y que hemos señalado y definido en los apartados anteriores: (1) redes de hiperenlaces, (2) temas y (3) actores.

- **Análisis de redes de hiperenlaces:** consiste en identificar los sitios web más relevantes relacionados con el tema de la transparencia y las relaciones que se dan entre ellos a través de hiperenlaces.

- **Análisis de temas:** consiste en identificar los diferentes subtemas o marcos de referencia existentes en la red temática a través del análisis de contenido de los textos de los sitios que forman la red.
- **Análisis de actores:** consiste en identificar los actores que forman parte de la web, su papel en la estructura de la red y las posturas que defienden respecto al tema de discusión.

IC6

El tercer objetivo de investigación o **Diseñar un sistema de análisis de redes temáticas en la web** se responde en la primera parte del **Capítulo 4** de esta tesis.

IC7

El análisis de redes temáticas debe desarrollarse en tres fases que corresponden al análisis de los tres componentes que forman parte de una red temática: (1) **análisis de redes de hiperenlaces**, (2) **análisis de temas** y (3) **análisis de actores**.

1.1.4. Cuarto objetivo: Aplicar el sistema de análisis al caso de la red temática de la transparencia en Chile.

El sistema previamente desarrollado se aplicará al estudio de la red temática de la transparencia en Chile con el fin de identificar la red de hiperenlaces relacionada con el tema elegido para el caso de estudio y los flujos de información que se producen en la red y las relaciones que construyen los actores que forman parte de ella. Una vez analizados los resultados se valorará la idoneidad del sistema de análisis en relación al objeto de estudio de esta investigación y su posible aplicación en otros casos.

IC8

El cuarto objetivo de investigación o **Aplicar el sistema de análisis al caso de estudio de la red temática de la transparencia en Chile** se responde en la segunda parte del **Capítulo 4** de esta tesis.

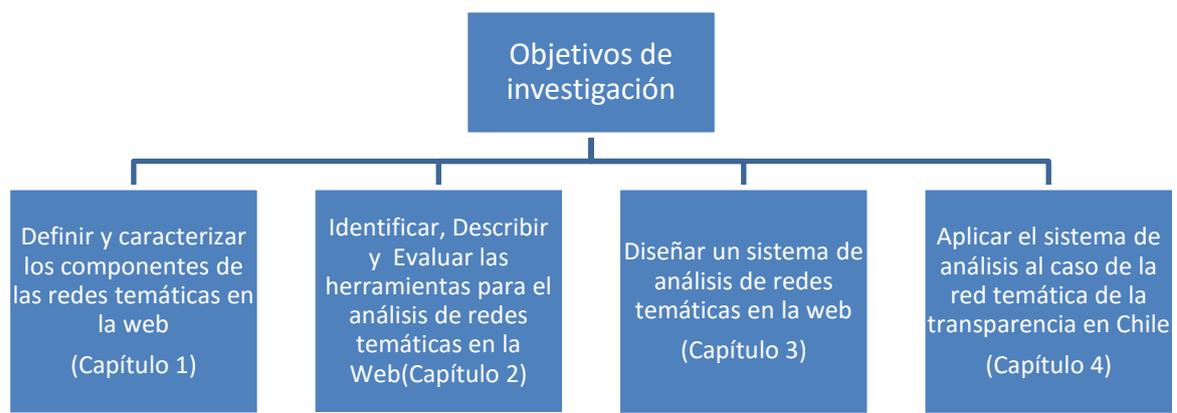


Figura 1. Resumen de los cuatro objetivos de investigación de Redes temáticas en la web: estudio de caso de la red temática de la transparencia en Chile.

1.2. Motivación y oportunidad de la investigación

El ser humano es un animal tecnológico; lo planteaba el gran Stanley Kubrick en su inmortal 2001 Odisea en el espacio y la transformación que está provocando el desarrollo de Internet en prácticamente todos los ámbitos de la sociedad parece confirmar esta idea al menos en el momento actual en el que vivimos. A lo largo de la historia de la humanidad se han producido innumerables avances tecnológicos pero solo unos pocos han sido capaces de transformar los sistemas de producción y las estructuras sociales en su conjunto. Como defiende Genís Roca, arqueólogo reconvertido en estudioso de lo que él denomina Sociedad Digital, una tecnología es relevante en la medida en que es capaz de transformar la forma en que alimentas a

los tuyos. Es decir, una tecnología es realmente transformadora cuando es capaz de transformar los sistemas de producción de una sociedad.

A lo largo de la historia ha habido innumerables avances tecnológicos, algunos de ellos muy importantes, pero solo unos pocos realmente disruptivos, transformadores de los sistemas de producción y de las estructuras sociales. Si preguntásemos a un historiador o a un investigador social cuáles han sido los avances tecnológicos más importantes de la historia de la humanidad, probablemente nos dibujaría una imagen similar a la que presentamos en la Figura 2.

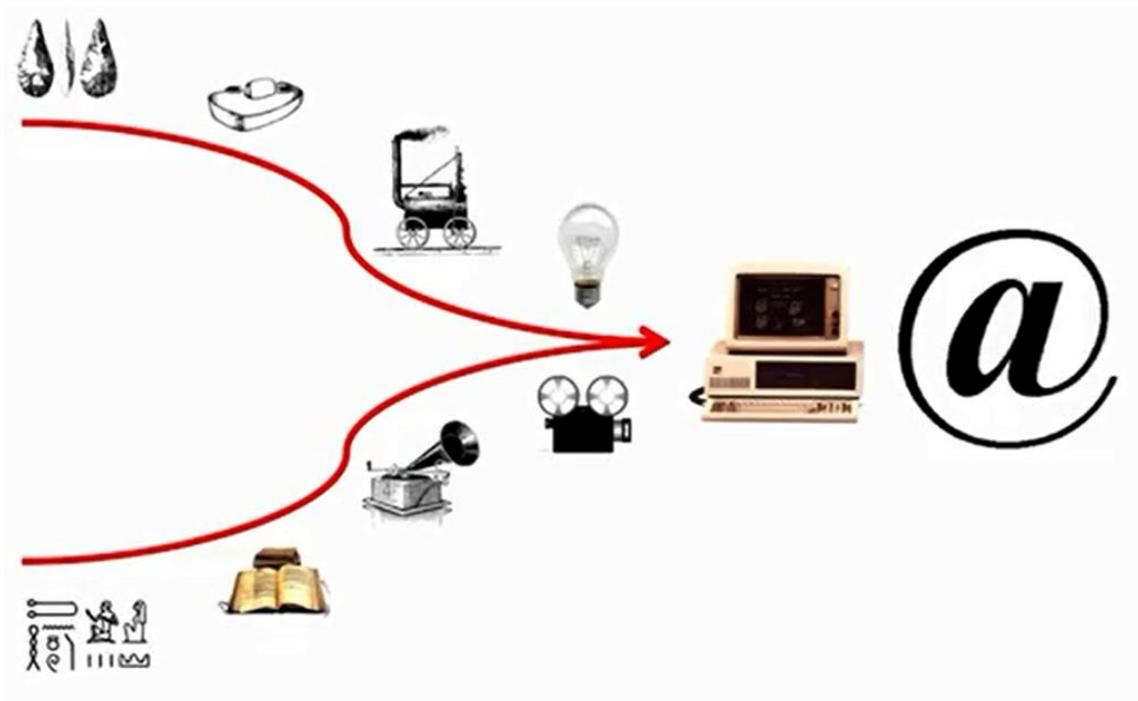


Figura 2. Representación de los avances tecnológicos más importantes de la historia de la humanidad. Extraído de Roca (2012).

En la Figura 2 vemos dos líneas separadas a lo largo de las cuales se van representando diferentes avances tecnológicos en un orden temporal a lo largo de la historia. En la línea superior se representan los avances tecnológicos transformadores de los sistemas de producción: las piedras talladas que nos permitieron cazar y procesar la carne, las piedras para moler que nos permitieron procesar el grano que ya

empezábamos a cultivar, la máquina de vapor que permitió el desarrollo generalizado de la revolución industrial, la electricidad que nos permitió iluminar las ciudades y no depender de los horarios naturales de luz para producir y consumir sin trabas.

En la línea inferior se representan los avances tecnológicos relacionados con la producción, gestión y consumo de la información, son las tan cacareadas tecnologías de la información que hoy en día, y erróneamente, asociamos únicamente a la tecnología digital. En esta línea inferior se representa en primer lugar la escritura, a través de unos jeroglíficos aparentemente egipcios, que nos permitieron por primera vez dejar constancia física de nuestra memoria, de nuestros acuerdos económicos y nuestras leyes. En segundo lugar encontramos el libro moderno, y a él asociamos la invención de la imprenta y los comienzos de la difusión masiva del conocimiento. En tercer lugar encontramos el gramófono, representando la primera tecnología de almacenamiento de información sonora. En cuarto lugar encontramos la cámara de vídeo, como ejemplo del nacimiento de la tecnología de producción y almacenamiento de información visual.

Si pensamos en estas tecnologías de la información y su relación con el sistema productivo, ese que nos permite alimentar a los nuestros, nos daremos cuenta de que ninguna de ellas se puede considerar influyente en este aspecto. Sin duda provocaron cambios importantes en la sociedad pero en ningún caso afectaron al trabajo de la mayoría de la población o transformaron la estructura de la sociedad. Más bien fueron avances que permitieron nuevas formas de difusión de la cultura que afectaron principalmente a grupos privilegiados de la sociedad.

Sin embargo, al llegar al punto donde se une la línea inferior y superior de la Figura 2, encontramos la representación de la tecnología digital en forma de un ordenador y posteriormente la representación del desarrollo de Internet a través del símbolo de la arroba. Por primera vez en la historia de la humanidad una tecnología de información es al mismo tiempo una tecnología transformadora de los sistemas de producción y con ellos de las estructuras sociales. Por primera vez, una tecnología de la información tiene la capacidad de influir en la forma en que la mayoría de la población desarrolla su trabajo y consigue el sustento para los suyos.

Internet ha resultado ser pues una tecnología disruptiva, es decir, una tecnología no solo capaz de transformar los sistema de información, sino que transformadora de

sistemas varios como el sistema de medios de comunicación, el sistema educativo, el sistema empresarial y en última instancia, aunque todavía de forma incipiente, del sistema político y de las relaciones entre la ciudadanía y las estructuras de poder.

En 2001, Luis Bassat escribió: “Con el nacimiento de internet, no solo ha nacido un nuevo canal de comunicación, ha nacido también un canal de investigación, un canal de venta y un canal de atención y de servicio al cliente. Con Internet ha nacido una nueva generación de consumidores; son aquellos críos que descubrieron la informática con el Spectrum y que ahora son consumados navegantes, que buscan, miran y compran utilizando su tarjeta de crédito. La televisión entró primero en la clase alta. Internet ha empezado por la gente joven”.

Cuando hablamos de internet, de lo digital, de los servicios de información en línea, no solo hablamos de sistemas de información, sino que además hablamos de sistemas de producción, y esto es lo que hace que las transformaciones que provoca afecten a la organización de la sociedad. Como señala Bauman en su obra *Tiempos líquidos* “Ahora la sociedad es una red y no una estructura: se percibe y se trata como una matriz de conexiones y desconexiones” o como señala Castells: “La sociedad red es la estructura social de nuestro tiempo. No es una sociedad del futuro compuesta por internautas solitarios [...]. Tampoco es la tierra prometida de las nuevas tecnologías [...]. Es, simplemente, la sociedad en la que hemos ido entrando [...] en un proceso de transición a partir de la sociedad industrial”.

Estas transformaciones sociales provocadas por la tecnología digital y el desarrollo de Internet, han provocado también cambios sustanciales en el sistema comunicativo y en los procesos de formación de la opinión pública. Los grandes medios de comunicación que funcionaban como filtro en nuestra interpretación de la realidad y que en relación con el poder político y empresarial, controlaban la agenda pública, han visto como su poder de control sobre la opinión pública se veía fracturado por la aparición de nuevas formas de publicación y acceso a la información. La agenda temática de la sociedad, que durante el siglo XX, había sido relativamente homogénea debido a que pocos medios de comunicación controlaban el sistema comunicativo, se convierte ahora en una agenda temática fragmentada. El poder de selección y acceso a la información de la ciudadanía ha aumentado debido a la facilidad para publicar y difundir información en formato digital utilizando plataformas sociales de comunicación. Lo que hace que

un individuo pueda crear su propia agenda temática a través del acceso a la información a través de sistemas de información personalizables.

Los datos de 2015 de la investigación en consumo de noticias que hace anualmente el Reuters Institute en colaboración con la Universidad de Oxford, nos dejan ver, por ejemplo, una aceleración del ritmo hacia el consumo de noticias a través de plataformas sociales. El papel desempeñado por Facebook y otras plataformas sociales que permiten la búsqueda, discusión e intercambio de noticias, están jugando un papel muy importante entre especialmente entre los grupos más jóvenes de la sociedad. Aunque la mayoría de noticias que consumimos, todavía provienen de medios tradicionales o conocidos, la forma de acceder a ellas está cambiando notablemente. El camino para llegar a una noticia, cada vez depende más de una segunda plataforma que sirve de puente al medio en que está alojada la noticia. Es decir, cada vez visitamos menos las páginas de inicio de los medios de comunicación, y nos dejamos guiar hasta las noticias de nuestro interés a través de buscadores de noticias o a través de recomendaciones en nuestras redes sociales y plataformas de comunicación. Lo vemos a continuación en la Figura 3.

	 UK	 US	 FRA	 GER	 IRE	 DEN	 FIN	 ITA	 SPA	 JPN	 BRA	 AUS
Direct to news brand	52%	36%	27%	26%	44%	54%	63%	20%	36%	15%	46%	33%
Search	32%	40%	40%	45%	46%	29%	26%	66%	54%	54%	52%	49%
Social Media	28%	35%	21%	20%	36%	38%	28%	33%	35%	14%	48%	41%
Email	10%	25%	21%	15%	9%	24%	9%	17%	14%	15%	23%	20%
Mobile notifications and alerts	10%	13%	14%	9%	9%	9%	7%	7%	8%	7%	11%	9%

Figura 3. Datos sobre la formas de acceso a las noticias ordenados por países y tipos de plataforma. Extraído de Reuters Institute for the Study of Journalism, (2015).

Si analizamos en detalle la Figura 3, nos daremos cuenta de que efectivamente, el porcentaje de la audiencia de los medios de comunicación que llega a las noticias a través de su página de inicio o que busca las noticias a través de un medio de comunicación en concreto, ha descendido en la mayoría de los países por debajo del 50 %. En países como Japón o Italia ese porcentaje se encuentra entre el 15 y el 20 %. Vemos sin embargo, que el porcentaje de personas que llega a las noticias a través de la búsqueda, en servicios como Google News o similares, ronda el 50 % en muchos

países y en todos los casos supera el 25 %. Las redes sociales como filtro de búsqueda de noticias no alcanza los porcentajes que presentan los buscadores de noticias pero en la mayoría de los países estudiados ronda ese 25 % llegando incluso al 48 % en el caso de Brasil. El correo electrónico sigue manteniendo, a pesar de lo que podríamos aventurar, una posición importante en el filtro de noticias. En la mayoría de los casos alcanza un porcentaje superior al 10 % y en el caso de Estados Unidos mantiene un sorprendente 25 %. Las notificaciones y alertas móviles irrumpen en este tipo de estudios con porcentajes cercanos al 10 %.

De la misma forma que los medios de comunicación tradicionales funcionaban como filtro en el consumo de información, cabría pensar que esa función la están desarrollando ahora las plataformas sociales y que existe, por tanto, el riesgo de que estas se conviertan en los agentes de control de la agenda temática de las audiencias. Sin embargo, la misma investigación del Reuters Institute for the Study of Journalism, aporta indicios para negar esta posibilidad. Según los datos del estudio de este año 2015, el 76 % de los usuarios de plataformas sociales y el 73 % de los usuarios de buscadores han afirmado que algunas veces o frecuentemente acceden a fuentes de información no habituales a las que no habrían accedido de no haber sido guiados a ellas por estos servicios.

A pesar de esto, no podemos negar que los medios tradicionales de comunicación y sus marcas siguen teniendo un lugar preponderante en el sistema de medios y noticias. El mismo estudio llega a la conclusión de que una noticia alojada en un medio de marca reconocida es más probablemente escogido por los usuarios que una noticia de un medio de marca no conocida. Según este estudio, el 37 % de los encuestados reconocen como factor importante la marca del medio en su selección de noticias en las plataformas sociales, mientras que en los buscadores, es la relevancia del titular en relación a la búsqueda realizada, el factor fundamental en la selección de la noticia. Pero no solo cambia la manera de acceder a las noticias por parte de los usuarios, los medios de comunicación tradicionales tienen que competir ahora con medios de comunicación digitales, nacidos posteriormente al desarrollo de internet, que pasan a jugar un papel importante en el ecosistema de medios. La pluralidad de medios también provoca el aumento de la pluralidad informativa. Es decir, el aumento en el número de medios de comunicación provoca un aumento en la variedad de la

información a la que podemos tener acceso⁴. Las agendas temáticas ya no son homogéneas, sino que cada persona tiene la posibilidad de crear su propia agenda temática. La agenda pública monolítica característica de la sociedad de masas en el siglo XX, se convierte en una agenda pública fragmentada en la que la cantidad de medios a los que acceden los ciudadanos y las variadas formas de selección de información hace difícil estudiar los procesos de formación de la opinión pública con las teorías tradicionales de los efectos de la comunicación.

Consideramos que el concepto de Redes Temáticas se adapta mejor a las características actuales del sistema comunicativo que las tradicionales formas de análisis de noticias que planteaban los estudiosos de la agenda setting por permitir estudiar fenómenos complejos en red y no centralizados pero justificaremos esta afirmación en el apartado dedicado a las consideraciones previas al concepto de redes temática.

Durante el curso 2009 – 2010 cursé el Máster de Estudios Avanzados en Comunicación Social de la Universitat Pompeu Fabra, donde tuve la oportunidad de cursar la asignatura *Opinión Pública* impartida por los doctores Jordi Pericot, Jordi Berrio y Carles Pont. En ella se desarrolló un recorrido por las principales teorías de los Efectos de la Comunicación, entre las cuales destacaron para mí, la Teoría Sistémica de Nicklas Luhman y la Teoría de la Agenda Setting de McCombs y Shaw.

El contexto de consumo de información y características del sistema comunicativo actual que hemos comentado previamente, provocó un interés especial por estas teorías y fue la motivación para desarrollar un primer trabajo que se convirtió en mi Trabajo Final de Máster dirigido por el profesor Joan Corbella y que llevaba por nombre “Metodología para el estudio de la opinión pública y la comunicación política en Internet” (Castillejo, 2011).

En este trabajo inicial defendíamos que para muchos tipos de investigación en ciencias sociales existe la necesidad de obtener información sobre la opinión pública y que Internet puede ayudar a satisfacer esta necesidad porque es una fuente de información fácilmente accesible. Intentábamos además plantear una metodología de

⁴ Para más información sobre la relación entre la pluralidad de medios de comunicación y el pluralismo informativo, ver el estudio de López García, (2006), “Comunicación en red y mutaciones de la esfera pública”. *Zer*, 20, 231-249.

análisis completa y coherente que para ello desarrollaba un análisis de las herramientas, métodos y teorías utilizadas en el análisis de la opinión pública en Internet.

Ya que la idea de opinión pública en que nos basábamos se fundamentaba en la organización temática de la misma, analizábamos las teorías clásicas de la agenda-setting y de la tematización pero planteábamos que dada la característica complejidad de la sociedad actual y la fragmentación de la agenda pública, se hacía necesario estudiar el nuevo sistema comunicativo desde perspectivas menos deterministas. Entre ellas estudiábamos la teoría de redes complejas como un posible punto de partida para analizar los fenómenos de emergencia temática y autoorganización desde una perspectiva más amplia que tuviera en cuenta los numerosos factores que determinan el funcionamiento del sistema comunicativo en la actualidad.

A principios del año 2011, y ya dentro del programa de Doctorado en Comunicación Social de la Universitat Pompeu Fabra, presenté un proyecto de tesis dirigido por el doctor Lluís Codina cuyo título provisional fue “Redes temáticas en la web: metodologías para el análisis en el ámbito de la Comunicación Social”, y cuya evaluación positiva nos permitió comenzar la investigación cuyos resultados presentamos ahora.

En este proyecto de tesis surgía ya el concepto de red temática, al que llegamos a través de las investigaciones de Richard Rogers y concretamente a través de su artículo “*Operating Issue Networks on the Web*”. En este trabajo epistemológico se defiende que la Web es una cultura de conocimiento distinta de la de los medios de comunicación y como tal requiere de nuevas teorías y sistemas de análisis.

Internet, como espacio comunicativo no jerárquico en el que es posible presentar y recibir información y opiniones sin depender de la selección y descripción que tradicionalmente llevaban a cabo los medios de comunicación, permite un consumo proactivo de la información que nos obliga a formar parte del sistema de comunicación como sujetos activos y no como los sujetos pasivos que proponía el modelo de sociedad de masas característico del siglo XX. La motivación principal para desarrollar esta investigación es comprender esta nueva realidad y colaborar en la creación de nuevas teorías y estudios que permitan entender cómo se desarrollan los procesos de comunicación digital en general y de formación de la opinión pública en particular.

Consideramos que esta investigación tiene un valor teórico en relación al desarrollo del concepto de red temática, en concreto en su aplicación en los estudios en la web; una utilidad práctica en cuanto caracteriza y evalúa una serie de herramientas y propone un sistema de análisis; y finalmente tiene cierta relevancia social y política en cuanto analiza el caso de la red temática de la transparencia en Chile y aporta cierto conocimiento de las relaciones que existen entre los actores más relevantes relacionados con el tema y la postura que mantienen respecto al mismo.

A. Valor teórico: En el planteamiento inicial del proyecto de tesis que comentábamos previamente se introducía como objeto de estudio central el concepto de red temática que defendemos como una nueva aproximación a los estudios de los efectos de la comunicación y a otros ámbitos de investigación en ciencias sociales. Entendemos que esta investigación desarrolla el concepto de red temática y abre nuevas vías para la investigación, no solo de los procesos de formación de la opinión pública en la web, sino que ayuda al desarrollo del ámbito de la comunicación digital y propone nuevas líneas de investigación en Comunicación Social.

Uno de nuestros principales objetivos ha sido ampliar y caracterizar en detalle el concepto de red temática que si bien había sido utilizado previamente, consideramos no había sido suficientemente descrito en su aplicación a los estudios de redes temáticas en la web. Para conseguirlo hemos trabajado en la caracterización detallada de los componentes que según autores como Rogers, componen las redes temáticas en la web, es decir: redes de hiperenlaces, temas y actores.

Entendemos que los conocimientos aportados en esta caracterización de componentes de las redes temáticas pueden ser útiles, siempre con las adaptaciones que convengan en cada caso, a investigaciones relativas tanto al ámbito de la sociología, las ciencias políticas como las ciencias empresariales. En todas ellas puede ser necesario entender las relaciones que se producen entre grupos de actores a través que comparten un mismo ámbito temático y establecen relaciones a través de redes de hiperenlace, ya sea en sus sitios web o en plataformas sociales de comunicación.

B. Utilidad práctica. El diseño y desarrollo de un sistema de análisis de redes temáticas en la web ha supuesto integrar un conjunto de técnicas de recuperación,

análisis y representación de la información entre las que se encuentran el análisis de redes de hiperenlaces, el análisis de redes sociales, el análisis de contenidos y la visualización gráfica de redes.

A través de estas técnicas ha sido posible operacionalizar el desarrollo teórico del concepto de red temática y sus componentes en un conjunto de fases y técnicas de análisis que pueden ser aplicables al ámbitos de estudio diversos como los fenómenos de opinión pública, los movimientos sociales, los estudios de mercado o los procesos de comunicación política en la web puesto que en todos ellos existen y es necesario analizar la actividad de grupos de actores que comparten un ámbito temático de interés común y establecen relaciones a través de la tecnología digital del hiperenlace.

En esta investigación hemos estudiado además las herramientas que consideramos más adecuadas para el análisis de redes de hiperenlaces, VOSON, IssueCrawler, SocSciBot y Webometric Analyst, describiendo para ello sus características y componentes y evaluando la calidad de aspectos generales y funcionalidades específicas de este tipo de herramientas. Consideramos que esta caracterización y evaluación puede ser valiosa para cualquier investigación que requiera extraer información sobre los hiperenlaces que se producen en la web en general o, en el caso de Webometric Analyst, incluso de plataformas sociales de comunicación.

C. Relevancia social y política. Aunque hemos planteado el estudio de un caso con un valor principalmente instrumental que nos permitiera validar el sistema de análisis y el modelo teórico de análisis de redes temáticas en la web, consideramos que el análisis de la red temática de la transparencia en Chile puede tener un valor social y político desde el momento en que aporta cierto conocimiento sobre las instituciones públicas y privadas de Chile que tienen alguna relación con el tema de discusión de la transparencia del estado, pero también de organismos privados, y las relaciones que existen entre los diferentes actores involucrados en este debate.

Gracias a los resultados del estudio de la red temática de la transparencia en Chile ha sido posible generar conocimiento sobre los actores más relevantes del debate de la transparencia en Chile, las comunidades y relaciones que existen entre actores gubernamentales y actores del tercer sector y las diferentes perspectivas que utiliza

cada actor en relación al tema de la transparencia y que hemos podido conocer a través del análisis de contenidos de sus sitios web. Consideramos que cualquier aporte en este sentido puede ser valioso para las propias organizaciones que participan en la red y para cualquier actor interesado en formar parte de ella. De igual manera, los resultados del estudio de caso pueden ser relevantes para cualquier investigador que necesite datos sobre las organizaciones que tienen relación con la transparencia o para investigadores de la transparencia en general.

A través del desarrollo de un análisis-test de una Red Temática, la Red Temática de la Transparencia en Chile, hemos presentado un caso que ilustra las posibilidades de nuestra investigación en general, y del sistema de análisis en particular en el ámbito de la Comunicación. El sistema propuesto, además de permitir un detallado análisis de los actores, temas e influencias recíprocas en redes temáticas, facilita su representación mediante diferentes formatos de representación visual de la información. Por todo ello, podría ser aprovechado por el Periodismo de Datos o por cualquier ámbito de conocimiento que requiera de técnicas de análisis y representación de la información, lo que consideramos tiene en sí mismo una relevancia política como herramienta de acceso a la información y por tanto como herramienta que favorece el conocimiento y la transparencia de fenómenos de formación de la opinión pública y los debates que intervienen en la toma de decisiones políticas.

1.3. Preguntas de Investigación

1.3.1. Preguntas genéricas

Como se ha mencionado en el primer apartado de este capítulo, los trabajos previos que han dado lugar a esta investigación (Castillejo, 2011; Castillejo, Codina y Pont, 2014) han permitido plantear las siguientes preguntas genéricas (PG) que busca responder esta investigación:

PG1: ¿Es posible definir de forma no ambigua y operativa a efectos de análisis e investigación el concepto de red temática en el ámbito de la web?

PG2: ¿Cuáles son las herramientas que podemos considerar más adecuadas para el análisis de redes temáticas en la web?

PG3: ¿Podemos desarrollar un sistema de análisis suficientemente escalable y ágil para adaptarse a la dinámica característica de los fenómenos de opinión que tienen lugar en la web?

PG4: ¿Es posible aplicar el concepto de redes temáticas al estudio de los fenómenos de opinión que tienen lugar en la web?

1.3.2. Preguntas específicas

Partiendo de las preguntas genéricas anteriores, se presentan las siguientes Preguntas Específicas (PE) articuladas alrededor de los cuatro ejes conceptuales básicos de este trabajo (**Definición y Caracterización de Componentes, Caracterización y Evaluación de Herramientas, Sistema de Análisis y Estudio de Caso**):

Eje de la Definición y caracterización de componentes de una red temática (primer objetivo de investigación)

PE1: ¿Es posible identificar de forma operativa los componentes principales y suficientes a efectos de análisis e investigación de una red temática en el ámbito de la web?

PE2: ¿Es posible presentar un conjunto de propuestas conceptuales sobre las características de los componentes que forman parte de una red temática en el ámbito de la web?

Eje de la Caracterización y Evaluación de Herramientas (segundo objetivo de investigación)

PE1: ¿Cuáles son las características de las herramientas que se pueden considerar más adecuadas para el análisis de redes temáticas en la web?

PE2: ¿Cuáles son los parámetros e indicadores que permiten evaluar las herramientas relacionadas con el análisis de redes temáticas en la web?

Eje del Sistema de Análisis (tercer objetivo de investigación)

PE1: ¿Qué fases y componentes puede tener un sistema de análisis escalable para analizar redes temáticas en la web?

PE2: ¿Cómo podrían operacionalizarse las tres fases del análisis de redes temáticas en la web: análisis de redes de hiperenlaces, análisis de temas y análisis de actores?

Eje del Caso de Estudio (cuarto objetivo de investigación)

PE1: ¿Es posible describir las características de la red de hiperenlaces que subyace a la Red Temática de la Transparencia en Chile?

PE2: ¿Es posible identificar las comunidades existentes en la Red Temática de la Transparencia en Chile?

PE3: ¿Es posible identificar las posiciones o marcos de referencia que defienden las diferentes comunidades existentes en la Red Temática de la Transparencia en Chile?

PE4: ¿Es posible identificar los actores más relevantes relacionados con el tema de la transparencia en Chile?

PE5: ¿Es posible identificar las relaciones que mantienen los actores más relevantes de la Red Temática de la Transparencia en Chile con el resto de actores de la red?

PE6: ¿Es posible identificar las relaciones que existen entre los actores más relevantes relacionados con el tema de la transparencia en Chile?

PE7: ¿Es posible identificar las posiciones que defienden los diferentes actores de la Red Temática de la Transparencia en Chile?

1.4. Marco Teórico

El marco teórico de esta investigación se desarrolla a través de los cuatro objetivos planteados y de los capítulos dedicados a cada uno de ellos. A continuación presentamos una síntesis de ellos con la que intentamos aportar una visión de conjunto de los diferentes marcos disciplinares estudiados y de las relaciones que establecemos entre ellos.

Recurrimos al marco disciplinar de la Sociología y concretamente al de la Teoría de Redes Sociales por lo que hace al concepto de Red Temática, incorporando conceptos de otras disciplinas como las Ciencias de la Información y (parcialmente) de las Ciencias de la Computación para poder trasladarlos al contexto de la Web y su estudio.

Igualmente, nos ha influido el marco teórico de la Comunicación Social, especialmente de las Teorías de los Efectos de la Comunicación, sobre todo como motivación para entender cómo podrían estudiarse ciertos fenómenos de opinión pública que tienen lugar en la web.

Otro marco teórico importante para nosotros es el de la dimensión sociológica del siglo XX, donde se consolida la Sociedad Industrial, se desarrollan los Medios de Comunicación de Masas y surgen los primeros intentos, en el ámbito de la Comunicación, por desarrollar teorías que permitan explicar el funcionamiento de los fenómenos de opinión pública. En el ámbito de la Sociología, surgen los primeros modelos sistémicos y se comienzan a aplicar conceptos de la Teoría Matemática de Redes para explicar el funcionamiento de los sistemas sociales dando lugar a la

Teoría de Redes Sociales. También, a principios del s. XXI empieza a tomar forma la Sociedad Red en la que se sitúa la Comunicación Digital y en la que cobran relevancia las teorías de Sistemas Complejos aplicadas al estudio de los fenómenos sociales y de opinión pública.

IC9

El Marco Teórico de esta tesis combina las **Ciencias de la Comunicación** (Teorías de los Efectos de la Comunicación y Opinión Pública) y la **Sociología** (Teoría de Redes Sociales) desde la Sociedad Industrial hasta la Sociedad Red.

A continuación señalamos cómo se desglosan los diversos marcos teóricos respecto a los distintos objetivos de investigación.

1.4.1. En relación al primer objetivo de investigación

En respuesta al primer objetivo de investigación o **Definir y caracterizar los componentes de las redes temáticas en la web**, que se desarrolla en el Capítulo 2, nos concentramos en identificar, caracterizar y definir cada uno de los componentes de una red temática, los cuales, como ya hemos señalado previamente, son: (1) las redes de hiperenlaces, (2) los temas y (3) los actores.

En este primer objetivo se recurre al marco teórico general de la Comunicación Social y de la Sociología. Además del doble marco teórico señalado se incorporan conceptos y perspectivas de otras disciplinas como las Ciencias de la Información, las Ciencias Políticas o las Ciencias Empresariales.

El marco teórico de la Comunicación Social, y concretamente las Teorías de los Efectos de la Comunicación que buscan explicar el funcionamiento de los procesos de formación de la opinión pública, nos sirve como acicate y motivación inicial para buscar una teoría que nos permita entender cómo funcionan esos procesos de formación de la opinión pública que tienen lugar o utilizan la Web como ecosistema de

comunicación. Respecto a este marco teórico de la Comunicación Social (CS) se tienen en cuenta especialmente:

- **CS1:** McCombs y Shaw, en su célebre artículo “*The Agenda Setting Function of Mass Media*”, definían la función de la agenda-setting como el resultado de la relación que se establece entre el énfasis manifestado por el tratamiento de un tema por parte de los *mass media* y las prioridades temáticas manifestadas por los miembros de una audiencia tras recibir los impactos de los medios de comunicación. Según esta hipótesis, cuanto mayor es el énfasis de los media sobre un tema, mayor es el incremento de la importancia que los miembros de una audiencia ofrecen a estos temas como orientadores de la atención pública. Existe, pues, una relación entre la agenda de los media y la agenda pública, siendo la primera la que inicia el proceso (McCombs y Shaw, 1972:177).
- **CS2:** La **Teoría de la Tematización** diferencia entre tema y tema político o *issue*; el tema no es sino una forma de generalización simbólica de situaciones singulares que posibilita la comunicación, mientras que un *issue* o tema político se forma a través de la consolidación de las posturas defendidas en una controversia generada en torno a un tema. Por tanto, los temas políticos o *issues* son objetos de discusión que se desarrollan a través de la interacción entre actores sociales y políticos con distintas posiciones, Marletti (1985).

La teoría de la tematización describe el sistemas político y el sistema comunicativo como un sistema complejo en el que, si bien los medios de comunicación tienen un papel determinante para la formación de la opinión pública, no se excluyen diferentes tipos de relaciones e influencias entre los distintos subsistemas. En palabras de Agostini, la tematización es un proceso que se realiza en la relación establecida entre el sistema político y la opinión pública, a través de la mediación de los medios de comunicación de masas. De esta manera los medios de comunicación han sido considerados no como los protagonistas sino sólo como mediadores de esta relación” (Agostini, 1984:531).

- **CS3:** El **análisis de redes de hiperenlaces**, desde la perspectiva de la Comunicación, defiende que en una red de hiperenlaces los actores son las organizaciones o los individuos que editan o administran los sitios web y los hiperenlaces son un mecanismo de asociación entre creadores de información.

En este tipo de redes, los hiperenlaces no son meros mecanismos para otorgar credibilidad a la manera de las citas bibliográficas, sino que, tal y como señalan Ackland y Gibson (2005), tienen al menos cinco funciones adicionales muy importantes como la provisión de información, la construcción de redes e identidad, el intercambio de audiencia y la amplificación de los mensajes. Por tanto, los hiperenlaces además de servir al flujo de información y la transmisión de mensajes, tienen una función representativa que influye en los procesos identitarios de los actores que los realizan y permiten construir lazos de comunidad (Shumate, 2012).



Figura 4. Resumen del Marco Teórico de la Comunicación Social en el 1º objetivo de investigación.

Respecto al marco teórico de la Sociología se tienen en cuenta:

- **S1: Luhman**, en su Teoría de Sistemas (1974) desarrolla el concepto de tema, según el cual, la opinión pública debe ser concebida como estructura temática de la comunicación pública. Para Luhman, de la transformación estructural del sistema político se ha derivado la creciente centralidad de los temas y de las

controversias políticas (*issues*), y por consiguiente, los temas se han convertido en recursos estratégicos esenciales de la política en las sociedades complejas. En el excelente análisis del modelo teórico de la tematización del profesor Badía, *De la persuasión a la tematización. Introducción a la comunicación política moderna*, encontramos una explicación certera de las características del proceso de tematización en la opinión pública, (1992:172).

Luhmann asigna a la opinión pública la función de mecanismo guía del sistema político que reduce la elevada contingencia de aquello que es políticamente y jurídicamente posible por la vía estratégica de adaptar la estructura temática de los procesos de comunicación pública a las variables exigencias decisorias de la sociedad y de su sistema político.

De la transformación estructural del sistema político, se ha derivado la creciente centralidad de los temas y de las controversias políticas (*issues*) por lo que concierne tanto a la orientación de los procesos decisorios, como al funcionamiento de las mismas lógicas de consenso de las sociedades contemporáneas. Por consiguiente los temas se han convertido en recursos estratégicos esenciales de la política en las sociedades complejas, lo cual ha de ser necesariamente planteado en el marco de referencia de las interacciones entre el sistema político y es sistema comunicativo.

La función de tematización remite finalmente a una dinámica evolutiva de diferenciación funcional de los distintos modelos profesionales de periodismo y de los distintos roles informativos ejercidos por los medios de comunicación.

- **S2: La Teoría de la Sociedad Red.** Una sociedad red es aquella cuya estructura social está compuesta de redes activadas por tecnologías digitales de la comunicación y la información basadas en la microelectrónica. Entendiendo por estructura social aquellos acuerdos organizativos humanos en relación con la producción, el consumo, la reproducción, la experiencia y el poder expresados mediante una comunicación significativa codificada por la cultura (Castells, 2009).

La descripción de la opinión pública que se hacía durante el siglo XX se basaba según Sampedro en la tríada que componen “medios, sondeos y urnas”, tres instituciones que hasta finales del siglo XX gestionaban la opinión pública y que

ahora se ven cuestionadas por las nuevas características y posibilidades del sistema comunicativo de la Sociedad Red.

La opinión pública contemporánea surgió a comienzos del siglo xx, con el entramado institucional que tejieron entre sí los medios de comunicación de masas y la popularización de las encuestas, al hilo de la extensión del sufragio universal. Medios, sondeos y urnas manifiestan "la opinión y las predisposiciones de la gente común, que son tenidas en cuenta (o debieran serlo) por los que ejercen el poder (o quieren ejercerlo) en público" (Sampedro, 2010: 141-142).

- **S3: Teoría de Redes Sociales.** Las redes de comunicación son las pautas de contacto creadas por el flujo de mensajes entre distintos comunicadores en el tiempo y en el espacio (Monge y Contractor, 2003: 39). La teoría de redes señaló que el patrón de relaciones que se establece en la web tiene influencia sobre el comportamiento de los actores que establecen un hiperenlace (Freeman, 2004).

De la aplicación de este concepto surge una nueva perspectiva en el estudio del hiperenlace: el análisis de redes de hiperenlaces (Park, 2003). Este tipo de análisis se ha aplicado al estudio de regiones de Internet tales como la esfera política de los Estados Unidos (Foot y Schneider, 2002), la visibilidad de la nanotecnología en Internet (Ackland et. al., 2007), o los sitios web de los políticos en Corea del Sur (Park y Thelwall, 2008).



Figura 5. Resumen del Marco Teórico de la Sociología en el 1º objetivo de investigación o Describir y caracterizar los componentes de las redes temáticas en la Web.

1.4.2. En relación al segundo objetivo de investigación

En respuesta al segundo objetivo de investigación o **Identificar, Describir y Evaluar las herramientas para el análisis de redes temáticas en la Web**, que se desarrolla en el Capítulo 3, recurrimos al marco teórico general de las Ciencias de la Computación, en primer lugar a la disciplina de la **Recuperación de Información (RI)**, concretamente al ámbito de los web crawlers, y en segundo lugar a la disciplina relacionada con la **Evaluación de Calidad de Software (ECS)**, teniendo en cuenta lo siguientes puntos en cada una de ellas:

- **RI1:** La Recuperación de Información (RI) es un campo de estudio que tiene por objetivo encontrar contenidos (normalmente documentos), de naturaleza no estructurada (normalmente texto) que satisfacen una necesidad de información entre una gran colección de contenidos (normalmente almacenada en ordenadores). Los web crawlers nacen en el ámbito de la Ingeniería Informática, vinculados con el área de la Recuperación de Información (*Information Retrieval*).

- **RI2:** Con el nacimiento de la Web, se desarrolla un área especializada vinculada con la Recuperación de Información dedicada al desarrollo de técnicas y herramientas para el estudio de la web: la Minería Web. Esta disciplina toma prestadas técnicas y metodologías del ámbito de la Minería de Datos, de ahí que adopte el nombre de Minería Web o *Web Mining*, y tiene por objetivo descubrir información útil a partir de la estructura de enlaces de la web, los contenidos de las páginas que la componen y los datos de uso de las mismas.
- **RI3:** Los web crawlers son programas que buscan, recuperan, analizan y en ocasiones almacenan los documentos de la Web. De la misma manera que un usuario puede navegar de una página web a otra a través de los hiperenlaces, los web crawlers utilizan los hiperenlaces para moverse a través de la Web visitando páginas y analizando la información que contienen para después almacenarla o indexarla en un índice o repositorio.

En relación a la **Evaluación de Calidad de Software:**

- **ECS1:** Las diferentes metodologías y orientaciones para diseñar el sistema de evaluación de software. A nuestro parecer, y apoyándonos en una revisión bibliográfica sistemática, son tres las perspectivas de la evaluación de calidad en software que más desarrollo están teniendo en el ámbito teórico y que más se están aplicando en el sector productivo: (1) software como producto, según el cual la calidad del software se puede evaluar a través de un conjunto predefinido de características (2) software como proceso, según la cual es posible mejorar la calidad del producto software a través de la aplicación de modelos de calidad que controlen las diferentes fases del proceso de desarrollo de un software, (3) software como modelo, que surge de la aparición y desarrollo de la Ingeniería dirigida por modelos y que basa el control de la calidad en la arquitectura del software y no en su código como tradicionalmente se había hecho.
- **ECS2:** Los diferentes modelos de calidad para la evaluación de software y las posibles características y subcaracterísticas de calidad frente a las cuales evaluar el software. Prestamos especial atención a los modelos de calidad jerárquicos basados en la perspectiva del software como producto para lo que

estudiamos más o menos en detalle los modelos de Los modelos de McCall, Bohem y las normas ISO/IEC 9126 y ISO/IEC 25000.

- **ECS3:** Los diferentes sistemas de medición para desarrollar las métricas de calidad de del software prestando especial atención a las propuestas de la norma y ISO/IEC 25000 y concretamente de la sección ISO/IEC 25022 - *Measurement of quality in use* que define específicamente las métricas para realizar la medición de la calidad en uso del producto.

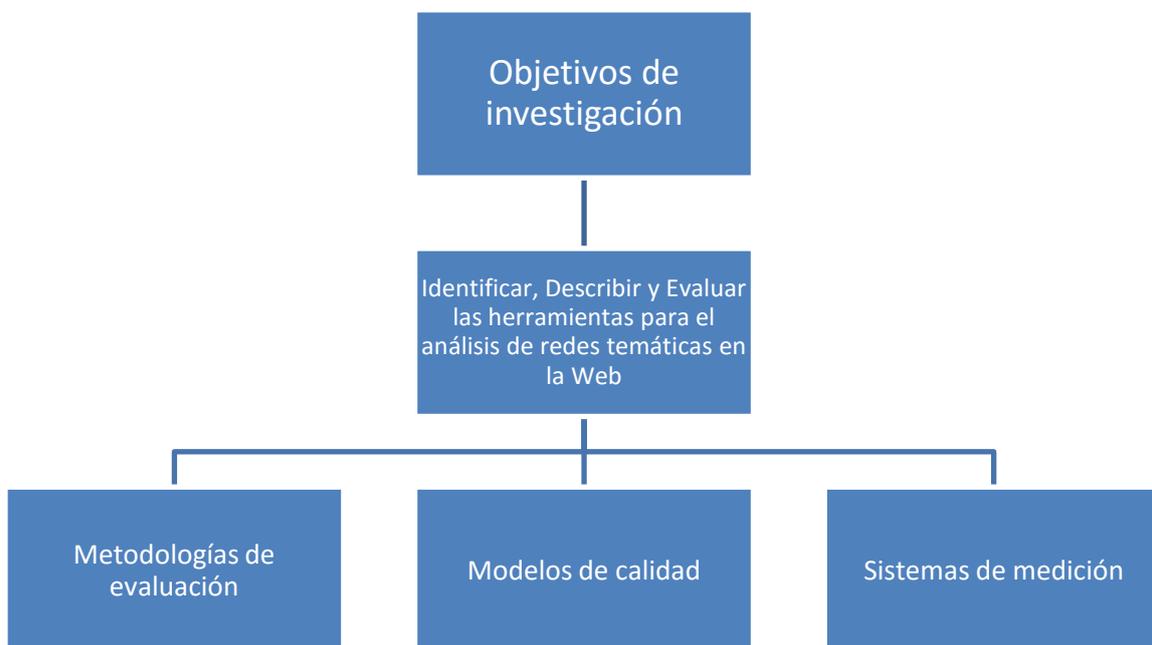


Figura 6. Resumen del Marco Teórico de la Evaluación de Calidad Informática en el 2º objetivo de investigación o Identificar, Describir y Evaluar las herramientas para el análisis de redes temáticas en la Web.

1.4.3. En relación al tercer objetivo de investigación

En respuesta al tercer objetivo investigación o **Diseñar un sistema de análisis de redes temáticas en la Web**, que se desarrolla en el Capítulo 4, recurrimos nuevamente a un doble marco teórico que tiene en cuenta las Ciencias de la Información y las Ciencias de la Comunicación. Del primero estudiamos la tradición del análisis de citas bibliográficas proveniente de la Bibliotecología y la Documentación para hacer un recorrido desde ella hasta el análisis de redes de hiperenlaces pasando

por la **cienciometría**, el análisis de redes de colaboración científica y su posterior aplicación y desarrollo en la web a través de la **webometría**.

- **CI1:** En la tradición de la **Bibliotecología y la Documentación** una cita bibliográfica implica que el autor que realiza la cita atribuye cierta relevancia al autor del documento citado. Esta idea es aplicada y adaptada al estudio de la actividad científica por la **cienciometría**, al estudio de información en la web por la **webometría** y al análisis de redes temáticas en esta investigación.
- **CI2:** La **cienciometría** surge como una rama especializada de la Bibliotecología y la Documentación que se dedica al análisis de la medición de la actividad científica y de ella nos interesan la teoría y la metodología relacionada con el análisis redes de colaboración científica. A partir de concepciones y teorías propias de la bibliometría y la **cienciometría**, la **webometría** lleva a cabo estudios informétricos sobre la información contenida en la Web. Según Bjernebern y Ingwersen (2004: 1217) la **webometría** es “el estudio de los aspectos cuantitativos que intervienen en la construcción y uso de recursos de información, estructuras y tecnologías que componen la web desde perspectivas bibliométricas e informétricas”.
- **CI3:** Desde la perspectiva de las Ciencias de la Información, y concretamente de la **webometría**, un hiperenlace actúa de forma similar a una cita bibliográfica ya que implica que el autor que realiza el enlace atribuye cierta relevancia a la página de destino. De esta forma, se pueden explicar las redes creadas a través de hiperenlaces en términos de credibilidad, prestigio y confianza (Kleinberg, 1999) de la misma manera que se hace con las citas bibliográficas en los estudios de **cienciometría**. En palabras de Thelwall (2012, la **webometría** “es el estudio de información contenida en la web a través de métodos principalmente cuantitativos para lograr objetivos de investigación en el ámbito de las ciencias sociales utilizando técnicas que no son específicas de un campo de estudio”.

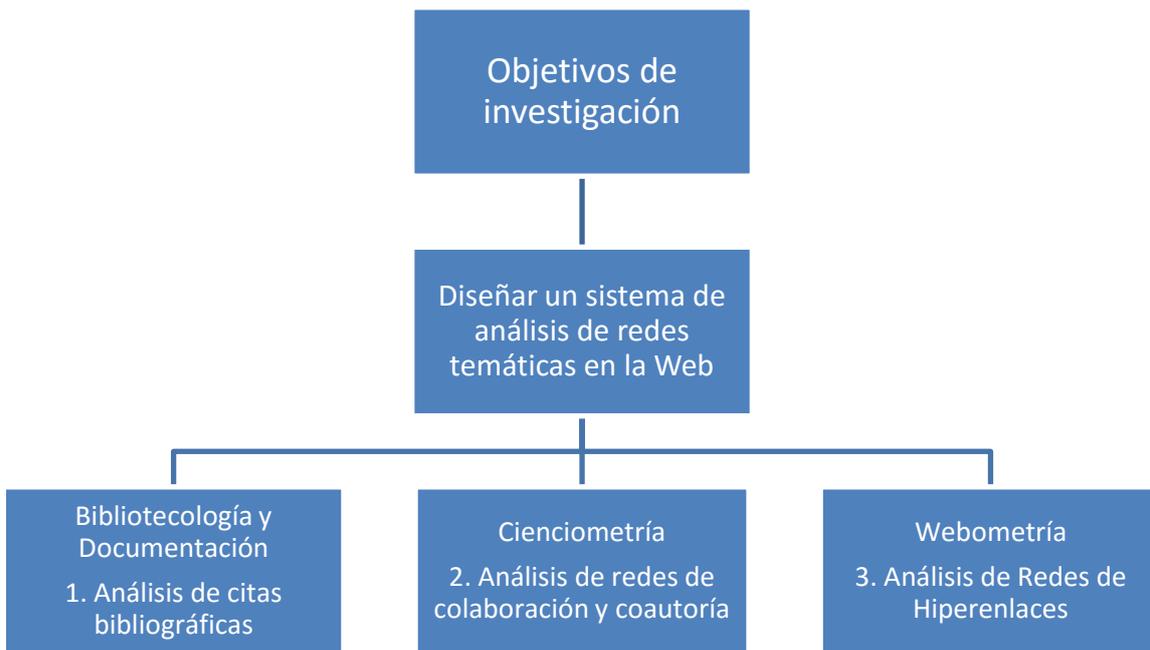


Figura 7. Resumen del Marco Teórico de las Ciencias de la Información en el 3º objetivo de investigación o Diseñar un sistema de análisis de redes temáticas en la Web.

Desde el ámbito de la Comunicación Social, buscamos un sistema de análisis que nos permita analizar los procesos de formación de la opinión pública que tienen lugar o utilizan la Web como ecosistema de comunicación. Para ello prestamos especial atención a la teoría relacionada con el análisis de redes de hiperenlaces y a la teoría de redes temáticas.

- **CS1:** Desde la perspectiva de la Comunicación Social, en una red de hiperenlaces los actores son las organizaciones o los individuos que editan o administran los sitios web y los hiperenlaces son un mecanismo de asociación entre creadores de información. En este tipo de redes, los hiperenlaces no son meros mecanismos para otorgar credibilidad a la manera de las citas bibliográficas, sino que, tal y como señalan Ackland y Gibson (2005), tienen al menos cinco funciones adicionales muy importantes como la provisión de información, la construcción de redes e identidad, el intercambio de audiencia y la amplificación de los mensajes. Por tanto, los hiperenlaces además de servir al flujo de información y la transmisión de mensajes, tienen una función representativa que influye en los procesos identitarios de los actores que los realizan y permiten construir lazos de comunidad (Shumate, 2012).

- **CS2:** Rogers (2002) define el concepto de red temática como una red de hiperenlaces por la que fluye información relacionada con una palabra clave. Como vemos, plantea un estudio del sistema comunicativo en Internet centrado en los circuitos de comunicación y en los flujos de información. Y propone además diferenciar los diferentes circuitos en función de palabras clave y organizarlos, por tanto, en torno a temas. Según este planteamiento, las redes de hiperenlaces en la web se construyen y se organizan en torno a temas de discusión, es decir, en torno a temas que generan cierta discrepancia o debate en la sociedad. Por ello, y para poder entender los procesos de formación de estas redes debemos prestar atención a los temas de discusión que provocan la construcción de estas redes temáticas.

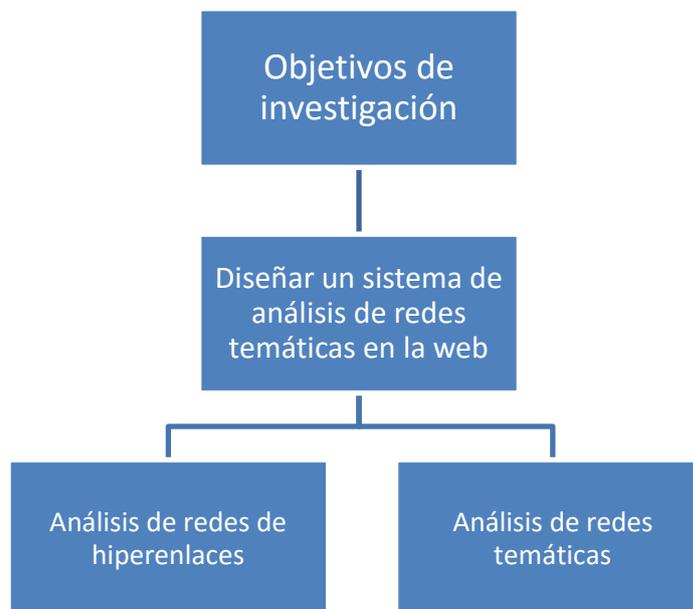


Figura 8. Resumen del Marco Teórico de la Comunicación Social en el 3º objetivo de investigación o Diseñar un sistema de análisis de redes temáticas en la Web.

1.4.4. En relación al cuarto objetivo de investigación

En respuesta al cuarto objetivo de investigación o **Aplicar el sistema de análisis al estudio de caso de la red temática de la transparencia en Chile**, que se desarrolla

tras el Diseño del Sistema de Análisis en el Capítulo 5, se realiza el estudio de caso basándonos en los marcos teóricos previos.

- De la **Teoría de Redes Temáticas en la Web** (primer objetivo de investigación), es decir, de la Comunicación Social con especial atención a la Teoría de la Agenda Setting, la Teoría de la Tematización y el Análisis de Redes de Hiperenlaces; y de la Sociología con especial atención en la Teoría de Sistemas de Luhman, la Teoría de la Sociedad Red de Castells y la Teoría de Redes Sociales.
- De la **Caracterización y Evaluación de Herramientas** para el Análisis de Redes Temáticas en la Web (segundo objetivo de investigación), es decir, de la Evaluación de Calidad en Software.
- Del **Diseño de un Sistema de Análisis** de Redes Temáticas en la Web (tercer objetivo de investigación), es decir, de las Ciencias de la Información, con especial atención en el análisis de citas bibliográficas, la cienciometría y la webometría; y de la Comunicación con especial atención en el análisis de redes de hiperenlaces y el análisis de redes temáticas.

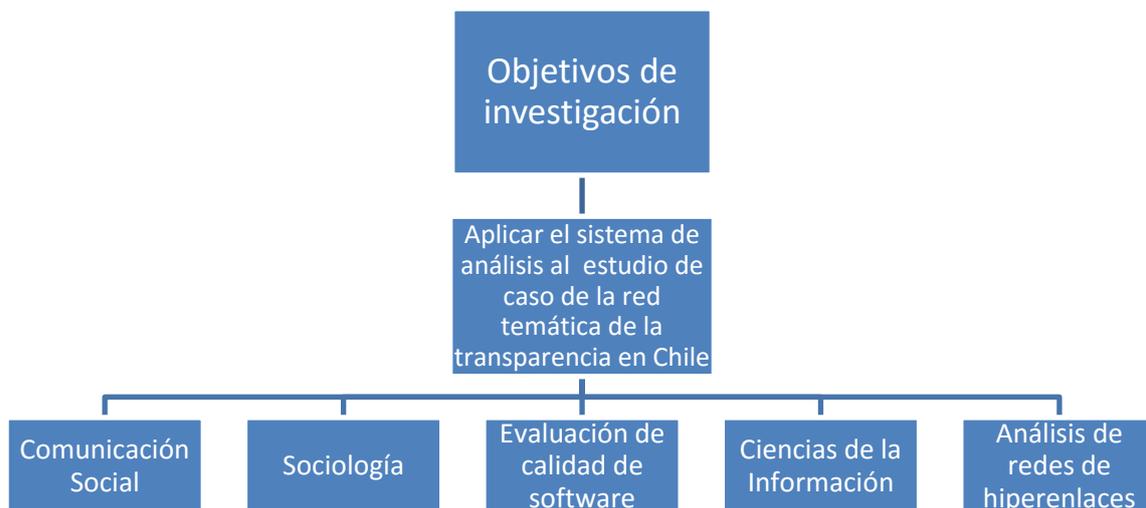


Figura 9. Resumen del Marco Teórico en el 4º objetivo de investigación o Aplicar el sistema de análisis al estudio de caso de la red temática de la transparencia en Chile.

Igualmente, este objetivo de investigación se desarrollará atendiendo a las características específicas del Caso de Estudio escogido, es decir, a los rasgos que hacen “única” la red temática de la transparencia en Chile. La selección de esta red se debe a que cumple con la definición de tema que propondremos en capítulos posteriores, es decir, es un conjunto de ideas, teorías o conceptos que generan controversia entre grupos de actores políticos, sociales o económicos.

Concretamente, hemos elegido el tema de la transparencia porque cumple ciertos requerimientos que lo hace idóneo para un trabajo de tipo exploratorio como este: (1) es un tema de debate o de discusión real, (2) hemos elegido deliberadamente un tema de debate sobre el cual en principio parece existir un consenso y evitar así que la elección misma del tema de debate sea controvertida, (3) el tema de discusión se restringe a un ámbito geográfico concreto debido a la necesidad de aplicar y probar la metodología propuesta sobre un tema de discusión conocido por el investigador, lo que consideramos aporta mayor grado de fiabilidad a la evaluación de los resultados.

1.5. Marco metodológico

La investigación que desarrollamos en esta tesis combina la tipología exploratoria (objetivos 1 y 3) y la tipología descriptiva (objetivos 2 y 4) para dar respuesta a las preguntas que planteamos.

Se trata de una investigación exploratoria cuando nos enfrentamos al primer objetivo de investigación o **Definir y caracterizar los componentes de las redes temáticas en la web**, que se desarrolla en el Capítulo 2.

El concepto de red temática había sido definido previamente en el ámbito de las Ciencias Políticas, principalmente por Heclo (1978), y posteriormente utilizado en diversos estudios de comunicación en la web (Rogers, 2002; Foot y Schneider, 2002; Park y Thelwall, 2008). Consideramos que las redes temáticas que se desarrollan en la web tienen características diferentes de aquellas que definía Heclo y que las investigaciones que han aplicado el concepto de red temática a los estudios de información en la web no describían con suficiente claridad sus componentes y características y, por tanto, nuestro primer objetivo de investigación trata de dar

respuesta a esta necesidad definiendo y caracterizando sus componentes que nosotros consideramos son tres: redes de hiperenlaces, temas y actores.

Nuestra investigación es igualmente exploratoria cuando nos enfrentamos al tercer objetivo o **Diseñar un sistema de análisis de redes temáticas en la web**, que se desarrolla en el Capítulo 3.

Es cierto que existen precedentes en el estudio de redes temáticas en la web que ya proponían una serie de técnicas de análisis más o menos completas y ordenadas. Nuestra propuesta trata de colaborar precisamente en el desarrollo de este tipo de sistemas de análisis, aportando lo que creemos es un sistema completo, ordenado y coherente que nos permite analizar los tres componentes de las redes temáticas de forma independiente pero relacionada.

Utilizamos una tipología de investigación descriptiva al intentar dar respuesta al segundo objetivo de investigación o **Identificar, describir y evaluar las herramientas de análisis de redes temáticas en la web**, que se desarrolla en el Capítulo 2.

La investigación descriptiva “busca especificar propiedades, características y rasgos importantes de cualquier fenómeno que se analice” (Hernández Sampieri, R., 2006), en nuestro caso, el fenómeno a analizar son las herramientas de rastreo de hiperenlaces o crawlers. En esta parte de la investigación hemos seleccionado las cuatro herramientas que consideramos más adecuadas para el análisis de redes temáticas en la web, hemos estudiado sus componentes y hemos evaluando sus características generales como herramientas de software y sus funcionalidades específicas como herramientas de rastreo de hiperenlaces.

En el cuarto objetivo de la investigación o **Aplicar el sistema de análisis al estudio del caso de la red temática de la transparencia en Chile** utilizamos igualmente un tipo de investigación descriptiva. Como señala Sampieri (2006) este tipo de investigación también “Describe tendencias de un grupo o población”. Aunque las redes temáticas no siempre son, o no son necesariamente, una realidad formal que se constituye como tal de forma expresa y explícita; con frecuencia, son realidades implícitas que el análisis es capaz de identificar, definir y caracterizar.

En esta fase de la investigación tratamos de describir los tres componentes de una red temática que habíamos definido en la fase inicial, la red de hiperenlaces, los actores y los temas que conforman la red temática de la transparencia en Chile. Hacemos además una interpretación de los resultados obtenidos en el análisis de esta red por lo que podríamos decir que el enfoque utilizado en esta fase de la investigación es un enfoque descriptivo - interpretativo.

En cuanto al enfoque de la investigación, utilizamos un enfoque cualitativo en los cuatro objetivos que nos planteamos, sin embargo, en el sistema de análisis que proponemos algunas fases se basan en datos cuantitativos y nosotros interpretamos de forma cuantitativa algunas características de la Red Temática de la Transparencia en Chile como la centralidad de los actores en la red o el tamaño de las comunidades que existen en ella. No consideramos que nuestra investigación tenga por eso un enfoque mixto, sino que el enfoque cualitativo es el que nos permite responder a todas las preguntas que nos hacíamos aunque en ocasiones nos apoyemos en datos numéricos para ello.

Para explicar el diseño metodológico de nuestra investigación nos vamos a apoyar, como hicimos en el marco teórico, en los cuatro objetivos planteados y en los capítulos dedicados a cada uno de ellos.

1.3.1. En relación al primer objetivo de investigación

En relación al primer objetivo de investigación o **Describir y caracterizar los componentes de las redes temáticas en la Web**, hemos seguido el siguiente diseño metodológico.

En primer lugar hemos desarrollado una revisión sistemática de la literatura relacionada con las redes temáticas tanto en la web como fuera de ella. Siguiendo las interpretaciones de Lluís Codina (2015), una **revisión sistemática** se caracteriza por cuatro dimensiones muy bien determinadas:

- **Sistemática** significa que no es arbitraria: ni sesgada ni subjetiva, sino que, por el contrario, hemos examinado la mejor producción científica disponible utilizando las mejores fuentes de información.
- **Completa** que se han usado sistemas de información de los que se presume que facilitan el acceso al **grueso de la producción de calidad** de una disciplina y a nivel internacional y que no hemos descartado ni incluido nada sin seguir otros criterios que los que hemos hecho explícitos, lo que conecta con los dos siguientes puntos.
- **Explícita** implica que se dan a conocer tanto las fuentes utilizadas como los criterios de búsqueda y de selección y exclusión.
- **Reproducible** es en realidad una consecuencia de todo lo anterior: dado que hemos sido sistemático y explícitos, permitimos a otros investigadores comprobar nuestro trabajo y, si lo desean, seguir los pasos y contrastar nuestros resultados para determinar su exactitud o su grado de acierto

La revisión sistemática que hemos llevado a cabo para responder al primer objetivo de investigación se ha desarrollado sobre el doble marco teórico de la Comunicación Social y de la Sociología, incorporando conceptos y perspectivas de otras disciplinas como las Ciencias de la Información, las Ciencias Políticas o las Ciencias Empresariales.

La búsqueda de los trabajos que han sido objeto de revisión en la construcción del marco teórico de este primer objetivo se ha realizado utilizando bases de datos científicas de diferentes ámbitos. Web of Science, Scopus y Jstor en el ámbito de la Ciencias Sociales y Políticas, aunque de Web of Science también se han recuperado trabajos relacionados con otros ámbitos como el de las redes complejas en el caso de los artículos de Stogratz o Barabasi, por poner solo un par de ejemplos.

En el ámbito de la Comunicación, la principal base de datos utilizada o mejor dicho, el agregador de bases de datos utilizado ha sido EBSCO. En ambos casos, las palabras clave y ecuaciones de búsqueda utilizadas han sido las que presentamos en la Tabla 1:

“redes temáticas”	“issue network”
“redes de hiperenlaces”	“hyperlink networks”
“agenda setting”	“public opinión”
“tematización”	“web agenda setting”
“opinión pública”	“complex networks”
“redes complejas”	“complex systems”
“sistemas complejos”	“issue emergence”
“emergencia temática”	

Tabla 1. Palabras clave utilizadas en la revisión sistemática para la construcción del Marco Teórico del 1º objetivo de investigación o Describir y caracterizar los componentes de las redes temáticas en la Web

En ocasiones, los resultados obtenidos con estas palabras clave resultaban demasiado numerosos; en esos casos tratábamos de reducirlos creando una ecuación más compleja con palabras clave adicionales como “web” o “internet” o combinando varias palabras clave como mostramos a continuación en la Tabla

“redes temáticas” AND (web OR internet)	“issue network” AND (web OR internet)
“opinión pública” AND (web OR internet)	“public opinion” AND (web OR internet)
“emergencia temática” AND (web OR internet)	“issue emergence” AND (web OR internet)
“tematización” AND “opinión pública”	(“complex system” OR “complex network”) AND “public opinión”
(“redes complejas” OR “sistemas complejos”) AND “opinión pública”	

Tabla 2. Ecuaciones de búsqueda utilizadas en la revisión sistemática para la construcción del Marco Teórico del 1º objetivo de investigación o Describir y caracterizar los componentes de las redes temáticas en la Web.

Generalmente, las búsquedas se han hecho teniendo en cuenta únicamente el título, las palabras clave y el resumen de los trabajos. Solo en los casos en que los resultados eran nulos o muy escasos, la búsqueda se amplió al texto completo de los trabajos.

Los criterios de inclusión y exclusión de los trabajos a analizar han tenido en cuenta en primer lugar la relevancia en función de las palabras clave y ecuaciones de búsqueda

utilizadas, pero también otros criterios como por ejemplo, la lengua y fecha de publicación de los trabajos o la repercusión de los mismos en función de su número de citas. En ocasiones, los trabajos seleccionados nos han llevado a otros trabajos que no habían aparecido en nuestro proceso de búsqueda en bases de datos especializadas. El árbol de citas de los trabajos más relevantes para nuestra investigación también ha sido analizado para identificar otros trabajos relacionados. Esta técnica de búsqueda ha sido especialmente relevante en la construcción del marco teórico relacionado con la descripción y caracterización del concepto de tema y actor que incluimos en nuestra definición de redes temáticas en la web.

Una vez seleccionados los trabajos se ha procedido a analizarlos, en primer lugar a través de la lectura de su resumen, sus conclusiones y la metodología empleada. Cuando el resultado de ese análisis inicial era “positivo”, desarrollábamos un análisis más completo para posteriormente sintetizar las ideas principales a considerar en la construcción del marco teórico correspondiente, en este caso, Describir y caracterizar los componentes de las redes temáticas en la web.

1.3.2. En relación al segundo objetivo de investigación

En relación al segundo objetivo de investigación o **Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la Web**, que se desarrolla en el Capítulo 2, hemos seguido el siguiente diseño metodológico.

En primer lugar, y siguiendo la misma aproximación metodológica utilizada en el objetivo anterior, hemos desarrollado una revisión sistemática de la literatura relacionada con los web crawlers en el ámbito de las Ciencias de la Computación y su relación con diferentes ámbitos de conocimiento como las Ciencias de la Información o la Comunicación Social; posteriormente hemos llevado a cabo una revisión sistemática de la disciplina de la Evaluación de Calidad de Software, igualmente en el ámbito de las Ciencias de la Computación, todo ello con el objetivo construir un marco teórico que sirva de contexto al desarrollo y uso de este tipo de herramientas y de las aproximaciones al concepto de calidad en el software y su aplicación en diferentes modelos de evaluación.

La búsqueda de los trabajos que han sido objeto de revisión en la construcción del marco teórico de este segundo objetivo se ha realizado utilizando bases de datos científicas del ámbito de las Ciencias de la Computación, principalmente IEEE Xplore, y en algunos casos también Web of Science y Scopus. A continuación, en la Tabla 3, presentamos las palabras clave y ecuaciones de búsqueda utilizadas para la consecución del segundo objetivo o Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la web:

“recuperación de información”	“information retrieval”
“informetría”	“informetrics”
“webometría”	“webometrics”
“minería de datos”	“data mining”
“calidad software”	“software quality”
“modelos de calidad”	“quality models”

Tabla 3. Palabras clave utilizadas en la revisión sistemática para la construcción del Marco Teórico del 2º objetivo de investigación o Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la Web

En ocasiones, los resultados obtenidos con estas palabras clave resultaban demasiado numerosos; en esos casos hemos tratado de reducirlos creando una ecuación más compleja con palabras clave adicionales como “software”, “crawler” o “web” o combinando varias palabras clave como mostramos en la Tabla 4.

“recuperación de información” AND web	“information retrieval” AND web
“informetría”	“informetrics”
“webometría” AND (crawler OR spider OR robot)	“webometrics” AND (crawler OR spider OR robot)
“minería de datos” AND web	“data mining” AND web
“calidad software” AND evaluación	“software quality” AND evaluation
“modelos de calidad” AND software	“quality models” AND software

Tabla 4. Ecuaciones de búsqueda utilizadas en la revisión sistemática para la construcción del Marco Teórico del 1º objetivo de investigación o Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la Web.

Los criterios de inclusión y exclusión de los trabajos, igual que en el caso anterior, han tenido en cuenta la relevancia en función de las palabras clave y ecuaciones de búsqueda utilizadas, la lengua y fecha de publicación y el número de citas recibidas. En este caso también nos hemos apoyado en el árbol de citas de los trabajos más relevantes para identificar otros trabajos relacionados.

El proceso de análisis de los trabajos ha seguido la misma rutina que hemos señalado en el primer objetivo. Es decir, se ha procedido a una lectura de su resumen, sus conclusiones y la metodología empleada y cuando el resultado del análisis inicial resultaba “positivo”, hemos desarrollado un análisis en profundidad para posteriormente sintetizar las ideas principales a considerar en la construcción del marco teórico correspondiente, en este caso, Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la Web.

Posterior a la revisión sistemática de la literatura relacionada con el desarrollo y uso de los web crawlers y de la evaluación de la calidad en el software, hemos propuesto un modelo de evaluación de calidad del software específico para el tipo de herramienta en cuestión, es decir, los web crawlers. El modelo propuesto, que se divide en un modelo de calidad general que podría ser común para cualquier software y un modelo de calidad funcional específico para este tipo de herramienta en el que se han analizado en profundidad las características propias de cada herramienta.

La selección de las herramientas a evaluar se ha basado en la revisión de la literatura que se desarrolló en el marco teórico relacionado con el primer objetivo de investigación o Describir y caracterizar los componentes de las redes temáticas en la Web. Las herramientas seleccionadas son las utilizadas en las investigaciones que consideramos más relevantes en el ámbito del análisis de redes temáticas en la web, (Rogers, 2008; Thelwall 2004a, 2004b, 2006; Ackland, 2005).

El modelo de calidad propuesto se aplica desde la perspectiva del usuario, en este caso, del investigador en ciencias sociales y de nuestra experiencia con las herramientas seleccionadas. Se trata por tanto de una evaluación subjetiva y cualitativa que utiliza un enfoque descriptivo, interpretativo y crítico.

1.3.3. En relación al tercer objetivo de investigación

En relación al tercer objetivo de investigación o **Diseñar un sistema de análisis de redes temáticas en la Web**, que se desarrolla en el Capítulo 4, hemos desarrollado una revisión sistemática en el ámbito de las Ciencias de la Información, concretamente de la tradición de la tradición del análisis de citas bibliográficas proveniente de la Bibliotecología y la Documentación para hacer un recorrido desde ella hasta el análisis de redes de hiperenlaces pasando por la ciencia métrica, el análisis de redes de colaboración científica y su posterior aplicación y desarrollo en la web a través de la webometría. En el ámbito de la Comunicación Social, hemos revisado diferentes sistemas de análisis de los procesos de formación de la opinión pública que tienen lugar o utilizan la Web como ecosistema de comunicación, prestando especial atención a la teoría relacionada con el análisis de redes de hiperenlaces y a la teoría de redes temáticas.

La búsqueda de los trabajos que han sido objeto de revisión en la construcción del marco teórico de este tercer objetivo se ha realizado utilizando bases de datos científicas, principalmente SAGE Journals, y en algunos casos Web of Science y Scopus. A continuación, en la Tabla 5, presentamos las palabras clave y ecuaciones de búsqueda utilizadas para la consecución del segundo objetivo o Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la web:

“redes temáticas”	“issue network”
“redes de hiperenlaces”	“hyperlink networks”
“esfera web”	“web sphere”
“redes sociales”	“social networks”
“redes científicas”	“scientific networks”

Tabla 5. Palabras clave utilizadas en la revisión sistemática para la construcción del Marco Teórico del 3º objetivo de investigación o Diseñar un sistema de análisis de redes temáticas en la Web

Cuando los resultados obtenidos con estas palabras clave resultaban demasiado numerosos hemos tratado de reducirlos creando una ecuación más compleja con palabras clave adicionales como “análisis”, como mostramos a continuación en la Tabla 6.

“redes temáticas” AND análisis	“issue network” AND analysis
“redes de hiperenlaces” AND análisis	“hyperlink networks” AND analysis
“esfera web” AND análisis	“web sphere” AND analysis
“redes sociales” AND análisis	“social networks” AND analysis
“redes científicas” AND análisis	“scientific networks” AND analysis

Tabla 6. Ecuaciones de búsqueda utilizadas en la revisión sistemática para la construcción del Marco Teórico del 3º objetivo de investigación o Diseñar un sistema de análisis de redes temáticas en la Web.

Los criterios de inclusión y exclusión de los trabajos, igual que en los casos anteriores, han tenido en cuenta la relevancia en función de las palabras clave y ecuaciones de búsqueda utilizadas, la lengua y fecha de publicación y el número de citas recibidas. En este caso también nos hemos apoyado en el árbol de citas de los trabajos más relevantes para identificar otros trabajos relacionados.

El proceso de análisis de los trabajos ha seguido la misma rutina que habíamos señalado en los objetivos anteriores. Es decir, se ha procedido a una lectura de su resumen, sus conclusiones y la metodología empleada y cuando el resultado del análisis inicial resultaba “positivo”, hemos desarrollado un análisis en profundidad para posteriormente sintetizar las ideas principales a considerar en la construcción del marco teórico correspondiente, en este caso, Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la Web.

Posterior a la revisión sistemática de la literatura relacionada con los sistemas de análisis de diferentes tipos de redes en la web, hemos propuesto un sistema de análisis propio. Para ello nos hemos apoyado en la descripción, caracterización y evaluación de herramientas que habíamos llevado a cabo en el Capítulo 2 teniendo en cuenta las posibilidades que ofrecían las herramientas evaluadas.

El resultado del estudio combinado de los marcos teóricos señalados previamente y de las herramientas seleccionadas ha dado como resultado un **Sistema de Análisis de Redes Temáticas en la Web** que consta de las siguientes fases y componentes:

- **Análisis de redes de hiperenlaces**, para el que proponemos desarrollar una fase inicial de selección de semillas, posteriormente un análisis de los enlaces

presentes en la red, un análisis de las medidas de redes sociales y finalmente una fase de visualización de las redes de hiperenlaces en sus diferentes variantes.

- **Análisis de temas**, para el que proponemos desarrollar un análisis de contenido de las diferentes comunidades presentes en la red; posteriormente un análisis del contenido relacionado con los actores que participan en la red; y finalmente una fase de visualización con nubes de palabras de los análisis previos asociados a comunidades y actores.
- **Análisis de actores**, para el que proponemos desarrollar una fase inicial de identificación de los actores que participan en la red; posteriormente una fase de análisis de redes sociales de cada uno de los actores; y finalmente una fase de visualización en la que presentamos una representación gráfica de las redes de vecindario de cada actor.

1.3.4. En relación al cuarto objetivo de investigación

En relación al cuarto objetivo de investigación o **Aplicar el sistema de análisis al estudio de caso de la red temática de la transparencia en Chile** que se desarrolla en el Capítulo 5, hemos tenido en cuenta las revisiones sistemáticas llevadas a cabo en los tres capítulos anteriores, concretamente, la relacionada con la Teoría de Redes Temáticas en la Web (primer objetivo de investigación); de la Caracterización y Evaluación de Herramientas para el Análisis de Redes Temáticas en la Web (segundo objetivo de investigación); y del Diseño de un Sistema de Análisis de Redes Temáticas en la Web (tercer objetivo de investigación).

Tomando como referencia los marcos teóricos construidos en relación con estos tres ámbitos, hemos aplicado el sistema de análisis propuesto a la Red Temática de la Transparencia en Chile. La metodología utilizada para conseguir alcanzar este cuarto objetivo se explica detalladamente en el Capítulo 5 pero a continuación presentamos un cuadro resumen de sus fases y componentes.

Fases y componentes del análisis de la Red Temática de la Transparencia en Chile

Análisis de la red de hiperenlaces

- Selección de semillas
- Análisis de enlaces
- Análisis de redes sociales
- Visualización de redes

Análisis de actores

- Identificación de actores
- Análisis de redes sociales
- Visualización de redes de vecindario

Análisis de temas.

- Análisis de contenido de comunidades
- Análisis de contenido de actores
- Visualización de nubes de palabras

2. REDES TEMÁTICAS EN EL ÁMBITO DE LA COMUNICACIÓN SOCIAL

El objetivo principal de este capítulo consistente en **Definir y caracterizar los componentes de las redes temáticas en la web** y se plantea, concretamente, a través de dos apartados:

- **Consideraciones preliminares en torno a las redes temáticas**, como campo y objeto de estudio, que recuperamos y ampliamos de un trabajo anterior (Castillejo, 2011).
- **Propuesta de componentes y caracterización** de redes temáticas en la web: redes de hiperenlaces, temas y actores.

2.1. Consideraciones preliminares en torno a las redes temáticas

Técnicamente, Internet parece ser un espacio comunicativo no jerárquico que permite a todo el mundo presentar y recibir información y opiniones sin depender de la selección y descripción que tradicionalmente llevaban a cabo los medios de comunicación. El desarrollo de la tecnología digital junto con un diseño de Internet pensado para la descentralización donde la posibilidad de crear y publicar contenidos está al alcance de cualquiera nos empuja hacia un nuevo tipo “consumo” activo en el que los procesos de comunicación nos permite tomar decisiones respecto a los contenidos que consumimos y nos obliga a formar parte del sistema de comunicación como sujetos activos y no como los sujetos pasivos que proponía el modelo de sociedad de masas característico del siglo XX.

En esta línea, Castells (2009: 88) defiende que las tecnologías de la información y la comunicación han desarrollado un nuevo tipo de comunicación que está provocando una transformación en el sistema comunicativo y en los procesos de formación de la opinión pública. El espacio público que desde la aparición de la sociedad postindustrial se había concentrado en los medios de comunicación de masas se transforma en un sistema donde (casi) cualquier persona puede crear y difundir sus informaciones. Esto

ha configurado un nuevo sistema de “comunicación masiva individual” en el que, como decimos, prácticamente cualquier ciudadano puede, o bien generar un mensaje o bien recuperar un contenido de la web para redistribuirlo y alcanzar, en cualquiera de los dos casos, a una audiencia global.

A pesar de este cambio de paradigma hacia una esfera pública en red, no podemos creer que hayamos llegado al ideal de democracia deliberativa de Habermas, en el que todos tenemos la misma capacidad de crear y difundir opiniones. La cantidad de información existente es tan grande que sería imposible para un usuario cualquiera conseguir la información deseada sin un mapa que le guíe. Sería ingenuo pensar que no existen procesos de selección o que estos procesos se regulan de forma democrática y no por un grupo minoritario actores. Sin embargo, parece lógico pensar que una transformación de este tipo en el sistema comunicativo ha de provocar una serie de cambios en los procesos de formación de la opinión pública.

Según Víctor Sampedro, las definiciones que hemos hecho de la opinión pública a lo largo de la historia, desde la antigua Grecia a la incipiente Sociedad Red de este siglo XXI, pasando por la Sociedad de Masas del siglo XX, dependen de las tecnologías y las instituciones que sirven para percibirla, expresarla y/o cambiarla.

Internet comenzó por cuestionar el número de actores que estaban implicados en la formación de la opinión pública. Ha potenciado las dimensiones discursivas de una esfera pública hasta entonces gestionada por periodistas, encuestadores y políticos que acabaron por constituir un mismo entramado institucional (o con intereses compartidos). Y que en lugar de articular un juego de contrapoderes se ve plagado de intercambios simbióticos y favores recíprocos. Desde la implantación de la imprenta no asistíamos a una colusión de modelos de representación tan antagónicos. Antagonismo que ha sido incrementado por el poder de Internet para acelerar los procesos sociales y superar las limitaciones espaciales. Hemos de abordar el carácter transnacional o, más bien, «glocalizado» de la opinión pública que se está gestando. La definición de la ciudadanía también está en juego y la noción de redes cobra la centralidad en el nuevo paradigma de la opinión pública aún en ciernes (Sampedro, 2010: 141-142).

Para Sampedro (2000: 20), existen dos tipos de formas de opinión pública, cada una de ellas con una forma de funcionamiento diferenciada.

- **La opinión pública agregada** es un resultado, la suma de juicios individuales a través del voto y los sondeos.
- **La opinión pública discursiva**, en cambio, es el proceso de un agente colectivo, el pueblo, que conversa en ambientes formales o informales procesando experiencias propias, conocimientos e informaciones.

Así, la opinión pública agregada es un resultado donde el público que importa es la mayoría que suma voluntades ya determinadas, mientras que la opinión pública discursiva es un proceso, donde el público es un colectivo de voluntades individuales que deliberan entre sí y se condicionan mutuamente. En consecuencia, la opinión pública discursiva siempre es más amplia que la agregada, que prima las mayorías y los números sobre los argumentos.

Es interesante recordar lo que señalaba Sabbah en una fecha tan temprana como 1985, en referencia a la proliferación de canales de televisión que estaba provocando el desarrollo de la tecnología de la televisión por el cable:

Los nuevos medios de comunicación determinan una audiencia segmentada y diferenciada que, aunque masiva en cuanto a su número, ya no es de masas en cuanto a la simultaneidad y uniformidad del mensaje que recibe. Los nuevos medios de comunicación ya no son medios de comunicación de masas en el sentido tradicional de envío de un número limitado de mensajes a una audiencia de masas homogénea. Debido a la multiplicidad de mensajes y fuentes, la propia audiencia se ha vuelto más selectiva. La audiencia seleccionada tiende a elegir sus mensajes, por lo que profundiza su segmentación y mejora la relación individual entre emisor y receptor (Sabbah, 1985: 219).

En teoría, y tal como apuntan algunas investigaciones (López-García, 2006), un mayor número de medios de comunicación provoca que el pluralismo aumente cuantitativa y cualitativamente. Debido a la aparición de un público activo muy diferente al tradicional, la presunción de que existe una exposición generalizada a un contenido político relativamente uniforme, que ha fundamentado las tres principales teorías sobre

los efectos de la comunicación política, *agenda-setting*, espiral del silencio y teoría del cultivo, no puede seguir considerándose una certeza (Blumler, 2001: 206). En este nuevo escenario parece claro que las hegemónicas posiciones de influencia y control del sistema político y los medios de comunicación pasan por una etapa de profunda transformación provocada, al menos en parte, por los cambios en los hábitos de consumo de información.

Como apuntaban Chaffee y Metzger (2001: 370) “en un futuro próximo el tema de discusión puede pasar de ser “¿qué hacen las grandes compañías de medios con la gente?” a convertirse en “¿qué hace la gente con los medios?”. El hecho es que la abundancia de canales posibilitada por las actuales tecnologías de la información y el diseño descentralizado de Internet, que favorece la creación y publicación de contenidos, están provocando la aparición de un número indeterminado de públicos y la fragmentación de la agenda pública.

En 2001, Donald Shaw, coautor del célebre artículo que dio origen a la teoría de la *agenda setting*⁵, planteaba la siguiente pregunta:

¿Es posible que los investigadores sigan llevando a cabo estudios para medir los efectos de los medios de comunicación cuando existe una cantidad de fuentes por las que el ciudadano es influido tan grande que hace la cuantificación y el análisis casi inabarcable? (Shaw, Stevenson y Hamm, 2001: 7-8).

Esta investigación tiene el objetivo de contribuir a establecer métodos para analizar los fenómenos de opinión en este nuevo ecosistema de comunicación y consideramos que el concepto de redes temáticas aplicadas al análisis de la opinión pública en la web puede contribuir a alcanzarlo pues propone una teoría adaptada a las características del sistema comunicativo actual, condicionado por las características de la Sociedad Red, y porque va acompañado de un sistema de análisis que permite hacer frente a las dificultades que supone el crecimiento en el número de fuentes de información que influyen y participan en los procesos de formación de la opinión pública en el nuevo sistema de comunicación dominado por la Web.

⁵ “The *Agenda setting* Function of Mass Media” (McCombs y Shaw, 1972: 177).

El concepto de Redes Temáticas ha sido estudiado y aplicado a la investigación en ámbitos de conocimiento diverso. A continuación, y para tratar de facilitar la plena comprensión del concepto y sus aplicaciones, desarrollamos un recorrido analizando los antecedentes más importantes en el ámbito de la Comunicación, las Ciencias Políticas y la Sociología.

Con la aparición de los medios de comunicación de masas surge un sistema comunicativo caracterizado por la uniformidad de contenidos y la exposición generalizada a una agenda temática única. Diferentes teorías de los efectos de la comunicación trataron de explicar a lo largo del siglo XX cuáles eran los mecanismos subyacentes de este sistema. Entre los primeros intentos están los de Walter Lippman. En 1922 Lippmann, en su libro *Public Opinion*, explicaba la necesidad de utilizar esquemas mentales que nos permitieran establecer modelos simples para hacer frente a la complejidad existente en la realidad.

El entorno real resulta en conjunto excesivamente grande, complejo y fugaz para que podamos conocerlo de forma directa. No estamos capacitados para manejar tanta sutileza y variedad, ni para considerar un número tan elevado de permutaciones y combinaciones. En consecuencia, por mucho que debamos actuar en él, nos vemos en la necesidad de reconstruirlo en modelos más asequibles para poder manejarlo. Podría decirse que estos modelos son como mapas que nos guían a través del mundo (Lippmann, 1922: 33).

Esta idea de complejidad de la realidad propuesta por Lippmann se mantiene en la teoría del sociólogo alemán Niklas Luhmann, quien a finales de los años 60 plantea la aparición de una sociedad compleja provocada por la aceleración del desarrollo industrial y tecnológico de la sociedad y por el incremento de organizaciones económicas, políticas, sociales o culturales. En este modelo de sociedad compleja, la tematización surge como principio organizador del sistema comunicativo, el sistema político y la opinión pública. A la manera de Lippman y sus esquemas mentales, Luhmann propone la tematización como ese modelo asequible que nos permitiría aprehender la complejidad de la realidad.

La opinión pública debe ser concebida como estructura temática de la comunicación pública [...]. La opinión pública no consiste en la generalización del contenido de las opiniones individuales mediante fórmulas generales, aceptables

por cualquiera que tenga uso de razón, sino en la adaptación de la estructura de los temas del proceso de comunicación política a las necesidades decisionales de la sociedad y de su sistema político (Luhmann, 1978: 87).

Si bien las ideas principales que hacen surgir la teoría de la tematización provienen de la obra de Luhmann, fueron un grupo de investigadores italianos del ámbito de la Comunicación Social (Agostini, 1984; Marletti 1985) los que la desarrollaron más ampliamente y la aplicaron a los estudios de opinión pública.

Dentro del ámbito hispanohablante, algunos de los estudios más destacados sobre la función de la tematización son los de Saperas (1987), Berrio (2002) y Badía (1992). De este último extraemos los tres presupuestos que fundamentan la hipótesis sobre la función de la tematización (Badía, 1992: 172).

- El concepto de opinión pública entendida ésta como una estructura selectiva de temas de la comunicación política. Luhmann asigna a la opinión pública la función de mecanismo guía del sistema político que reduce la elevada contingencia de aquello que es políticamente y jurídicamente posible por la vía estratégica de adaptar la estructura temática de los procesos de comunicación pública a las variables exigencias decisorias de la sociedad y de su sistema político.
- De la transformación estructural del sistema político, se ha derivado la creciente centralidad de los temas y de las controversias políticas (*issues*) por lo que concierne tanto a la orientación de los procesos decisorios, como al funcionamiento de las mismas lógicas de consenso de las sociedades contemporáneas. Por consiguiente los temas se han convertido en recursos estratégicos esenciales de la política en las sociedades complejas, lo cual ha de ser necesariamente planteado en el marco de referencia de las interacciones entre el sistema político y el sistema comunicativo.
- La función de tematización remite finalmente a una dinámica evolutiva de diferenciación funcional de los distintos modelos profesionales de periodismo y de los distintos roles informativos ejercidos por los medios de comunicación.

Como bien resume Badía (1992: 172), Luhmann entendía la opinión pública como una estructura de temas que sirve de guía y reduce así la elevada complejidad de los procesos de comunicación pública relacionados con las decisiones de la sociedad y de su sistema político. De la transformación estructural del sistema político, se ha derivado la creciente centralidad de los temas y de las controversias políticas (*issues*) por lo que los temas se han convertido en recursos estratégicos esenciales de la política en las sociedades complejas. Los temas obtienen su centralidad a partir de la creciente complejidad del sistema social y de las relaciones entre el sistema político, el sistema comunicativo y la opinión pública.

Entre los aportes posteriores a la teoría de la tematización, Marletti desarrolla una diferenciación entre lo que se define como tema, controversia e *issue* o tema político. Según esta diferenciación, el tema no es sino una forma de generalización simbólica de situaciones singulares que posibilita la comunicación, mientras que un *issue* o tema político se forma a través de la controversia generada en torno a un tema. Por tanto, los *issues* se originan en el contexto de los procesos de interacción entre los actores sociales y políticos. Su introducción en la agenda política pasa por el reconocimiento público de su relevancia y, seguidamente, por la generación de una dinámica de tomas de posición en los ambientes políticos y sociales influyentes (Marletti, 1985). Estudiaremos este aspecto con mayor profundidad en el apartado: 1.2. Redes temáticas en la web: una propuesta de componentes y caracterización. En él caracterizaremos el concepto de tema que consideramos uno de los tres componentes de las Redes Temáticas en la Web.

La teoría de la tematización describe el sistemas político y el sistema comunicativo como un sistema complejo en el que, si bien los medios de comunicación tienen un papel determinante para la formación de la opinión pública, no se excluyen diferentes tipos de relaciones e influencias entre los distintos subsistemas. En palabras de Agostini:

La tematización es un proceso que se realiza en la relación establecida entre el sistema político y la opinión pública, a través de la mediación de los medios de comunicación de masas. De esta manera los medios de comunicación han sido considerados no como los protagonistas sino sólo como mediadores de esta relación” (Agostini, 1984: 531).

Las teorías de la tematización y la *agenda-setting* son cercanas en su objeto de estudio y autores como Neumann las consideraron incluso sinónimas (1979: 439). McCombs y Shaw, en “The *Agenda setting* Function of Mass Media”, definían la función de la agenda-setting como:

El resultado de la relación que se establece entre el énfasis manifestado por el tratamiento de un tema por parte de los medios de comunicación de masas y las prioridades temáticas manifestadas por los miembros de una audiencia tras recibir los impactos de los medios de comunicación. (McCombs y Shaw, 1972: 177)

Según esta hipótesis, cuanto mayor es el énfasis de los media sobre un tema, mayor es el incremento de la importancia que los miembros de una audiencia ofrecen a estos temas como orientadores de la atención pública. Existe, pues, una relación entre la agenda de los medios de comunicación y la agenda pública, siendo la primera la que inicia el proceso.

Ambas teorías coinciden en dos aspectos fundamentales: (1) la necesidad que tiene nuestra mente de simplificar la realidad para entenderla y (2) la función de los medios de comunicación como puente entre la realidad y las imágenes de nuestra mente. Sin embargo, como señala Saperas, presentan diferencias importantes en su fundamentación:

Resulta muy diferente del estudio del establecimiento de la agenda temática ya que se fundamenta en la sociología sistémica, su contextualización del proceso de tematización en el sistema político y en el sistema comunicativo y su vinculación con las transformaciones tecnológicas y políticas que estamos experimentando actualmente son desconocidas por la *Agenda setting* (Saperas, 1987: 92).

La relación que establece la teoría de la tematización entre el sistema político y la opinión pública y su fundamentación en la idea de un sistema social complejo le permite abordar de manera, podríamos decir natural, las variaciones que ha vivido el sistema comunicativo en los últimos años para dar explicación a los procesos de formación de la opinión pública en la actualidad.

En 1978, Hecló desarrolla la teoría de redes temáticas como reacción a la teoría de los triángulos de hierro (Mills, 1958; Lowi, 1967; Ripley & Franklin, 1975; Hamm, 1983), a

la que algunos autores hacen referencia como “triocracia” (Galambos, 1983) o “*subgovernments*” (Hayden, 2002), y que explicaba las dinámicas políticas a través de la relación entre tres tipos de actores: poder legislativo, poder ejecutivo y grupos de interés de la sociedad civil. En palabras de Hayden:

Los grupos subgubernamentales que componen los tres lados del triángulo de hierro son (1) burócratas gubernamentales en cargos ejecutivos, (2) un comité legislativo o una comisión en la que el poder legislativo ha delegado autoridad, y (3) un grupo de interés corporativo integrado. Los participantes se organizan en relaciones recíprocas que se refuerzan mutuamente entre los intereses regulados y los reguladores (Hayden, 2002: 479).⁶

El concepto de red temática de Heclo ampliaba esta perspectiva para incluir en la ecuación a otro tipo de actores.

Al fijarnos en los pocos que son poderosos, tendemos a olvidarnos de todos aquellos cuyas redes de influencia provocan y guían el ejercicio del poder. Esas redes, o lo que yo llamaré “redes temáticas” son especialmente relevantes en la cada vez más intrincada y confusa política social que se vienen desarrollando en los últimos años (Hecló, 1978: 448).

La teoría de redes temáticas de Hecló plantea la existencia de un gran número de actores con diferentes niveles de relación o dependencia con otros actores del entorno. Si los triángulos de hierro planteaban la estabilidad de los participantes y la concreción de los temas que se discuten, en el modelo de redes temáticas los participantes entran y salen constantemente de la red y los temas de discusión son más amplios y flexibles.

Según Schlozman y Tierney:

Una red temática es una red de políticos o actores relacionados con el ámbito político que, formando parte del gobierno o no, están relacionados por su interés, conocimiento y compromiso con un área temática (Schlozman y Tierney, 1986).

⁶ Traducción libre de Hayden, F. (2002). "Policymaking Network of the Iron-Triangle Subgovernment for Licensing Hazardous Waste Facilities". *Journal of Economic Issues* 36 (2): 479 <http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1007&context=cbafacpub>

Como vemos los actores no tienen que ser necesariamente políticos, sino que se considerará como actor en la red a cualquier participante relacionado de alguna manera con el ámbito temático en cuestión. El ámbito temático ocupa el lugar central en el modelo de redes temáticas. De la misma manera, los recursos o capacidad de decisión del actor no determinan su participación en la red, sino que el mero interés o conocimiento de un área temática son recursos suficientes para participar, aunque sea de manera limitada, en ella.

Junto con el modelo de los triángulos de hierro y el de redes temáticas encontramos el modelo de comunidades políticas propuesto por Rhodes y Marsh (1992: 197). De manera similar a lo que proponía en el modelo de los triángulos de hierro, las comunidades políticas tienen un número limitado de miembros. Las comunidades políticas excluyen a determinados actores de manera consciente y están dominadas por intereses económicos y profesionales principalmente. Los participantes de estas comunidades tienen todos ellos recursos para la negociación y por tanto existe un equilibrio de poder entre ellos. Esto provoca, según Rhodes y Marsh, que los resultados de estas interacciones o negociaciones sean persistentes en el tiempo.

En contraposición con el modelo de comunidades políticas, las redes temáticas acogen un gran número de actores con un nivel de recursos diferenciado y con un amplio abanico de intereses no siempre concretos. Algunos de los actores que intervienen no disponen de recursos para la negociación en el mismo nivel que otros y por tanto las interacciones que se dan son desiguales. Esto provoca que en ocasiones los actores con menos recursos vean complicada su capacidad real de participación en los procesos de toma de decisión.

Otro modelo utilizado para el análisis de las dinámicas políticas, especialmente en temáticas relacionadas con la innovación tecnológica y el sector productivo, es el de la triple hélice. Según este modelo, el análisis de los patrones de comunicación y las relaciones entre organizaciones del gobierno, la industria y la academia es fundamental para la comprensión de sistemas económicos basados en el conocimiento.

La evolución de los sistemas de innovación, y el conflicto actual sobre qué camino se debe tomar en las relaciones universidad-industria, se reflejan en los diferentes

arreglos institucionales de las relaciones universidad-industria-gobierno (Etzkowitz y Leydesdorff, 1995).

Las consecuencias del cambio de un modelo de negociación “cerrado” como el de las comunidades políticas o los triángulos de hierro, a un modelo “abierto” como el de redes temáticas son varias. Para Wilkinson, Lowe y Donaldson (2010: 331):

- La falta de continuidad en las decisiones políticas que resultan de las interacciones entre los participantes;
- Un mayor nivel de conflicto entre los participantes.
- Una mayor variedad en los temas en discusión al interior de la red.

Desde el ámbito de la sociología, y en relación con la contraposición entre los modelos cerrados de comunidades políticas y triángulos de hierro frente al de redes temáticas, Castells (2009: 25) señala que “el poder político es tan solo una dimensión del poder, ya que las relaciones de poder se construyen en una interacción compleja entre diversas esferas de la actividad social”.

Según Castells, la incorporación de nuevos actores a las dinámicas de poder que se señala en el modelo de redes temáticas se debe al desarrollo de las redes de comunicación que a su vez está determinado por factores tecnológicos.

La capacidad de las redes para introducir nuevos actores y nuevos contenidos en el proceso de organización social, con relativa independencia de los centros de poder, se incrementó a lo largo del tiempo con el cambio tecnológico y más concretamente con la evolución de las tecnologías de la comunicación (Castells, 2009: 48).

Como vemos, el modelo “abierto” de redes temáticas está más cerca de las ideas de Castells sobre las relaciones entre comunicación y poder que las del modelo “cerrado” de comunidades políticas o triángulos de hierro. Queremos recuperar algunas ideas sobre el concepto de Sociedad Red que propone Castells ya que consideramos que desde el ámbito de la sociología explica, respalda y justifica nuestra decisión de utilizar el concepto de redes temáticas para desarrollar un sistema de análisis de la opinión pública en el sistema sociocomunicativo actual.

Para Castells (1997: 549) “las funciones y los procesos dominantes en la era de la información cada vez se organizan más en torno a redes” y en estos procesos dominantes están incluidos los procesos de producción, los de consumo, los de reproducción, los de la experiencia y los de poder, expresándose todos ellos a través de una comunicación codificada por la cultura.

Según Castells (2009: 49), las redes se convirtieron en la forma organizativa más eficiente como resultado de tres rasgos fundamentales:

- **Flexibilidad:** las redes pueden reconfigurarse en función de los cambios en el entorno, manteniendo su objetivo aunque varíen sus componentes. Son capaces de soslayar los puntos de bloqueo en los canales de comunicación para encontrar nuevas conexiones.
- **Adaptabilidad:** pueden expandirse o reducir su tamaño con pocas alteraciones.
- **Capacidad de supervivencia:** al no poseer un centro y ser capaces de actuar dentro de una amplia gama de configuraciones, las redes pueden resistir ataques a sus nodos y a sus códigos, porque los códigos están contenidos en múltiples nodos que pueden reproducir las instrucciones y encontrar nuevas formas de actuar. Por ello, sólo la capacidad de destruir físicamente los puntos de conexión puede eliminar la red.

Como señalan Monge y Contractor: en el ámbito social, “las redes de comunicación son las pautas de contacto creadas por el flujo de mensajes entre distintos comunicadores en el tiempo y en el espacio” (2003: 39). En esta afirmación Monge y Contractor introducen el concepto de flujos que tal y como indica Castells:

El poder de los flujos tiene prioridad sobre los flujos de poder. La presencia o ausencia en la red y la dinámica de cada una frente al resto son fuentes cruciales de dominio y cambio en nuestra sociedad: una sociedad que, por la tanto, puede llamarse con propiedad la sociedad red, caracterizada por la preeminencia de la morfología social sobre la acción social. (Castells 1997: 149).

Esta idea de la superioridad de los flujos de información sobre el contenido de los mensajes y de la estructura social sobre la actividad social, será fundamental para

comprender el interés de aplicar el concepto de redes temáticas al análisis de la opinión pública en la web. Para entender los factores que determinan la emergencia de un tema en la esfera pública y teniendo en cuenta las propuestas de Lippman, Luhmann y Castells para entender el sistema comunicativo actual como un sistema complejo, consideramos importante estudiar el concepto de complejidad como característica del sistema social actual y su intrincada interrelación con el sistema comunicativo.

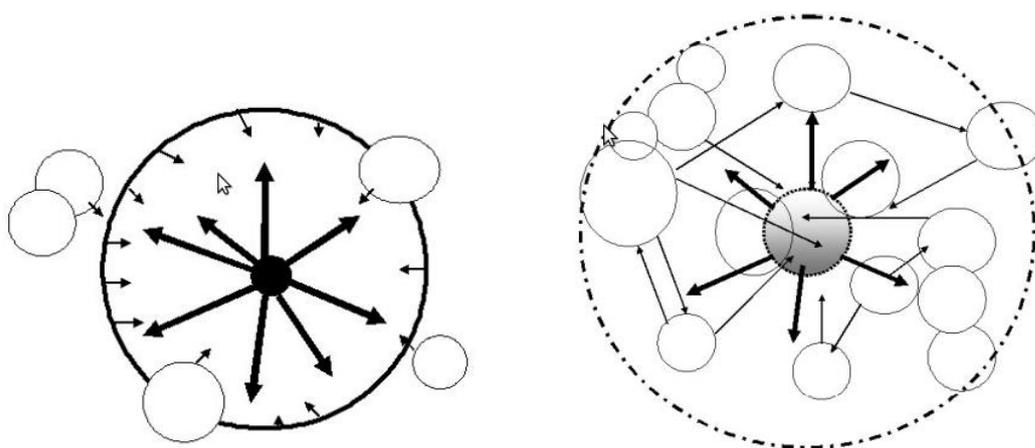


Figura 10: Modelo comparado de esfera pública en la sociedad de masas y en la sociedad red respectivamente. Extraído de López-García (2000: 236-240).

La teoría de la complejidad y de los sistemas complejos nace de la teoría de redes que tiene sus orígenes en el estudio de los puentes de Koninsberg por parte de Euler (1736), o más recientemente en la publicación del artículo de Erdos y Rényi (1959) sobre redes aleatorias. Estas teorías de redes que surgen en el ámbito de las Matemáticas, dieron lugar a la Ciencia de los Sistemas Complejos, que recientemente ha irrumpido con fuerza prácticamente en todas las ramas de la ciencia incluyendo entre ellas a las Ciencias Sociales en general y a la Comunicación Social en particular.

La Ciencia de los Sistemas Complejos, Ciencia de las Redes Complejas o Ciencia de las Redes, variantes todas ellas de un mismo ámbito de conocimiento, (Barabasi, 2002; Watts, 2003; Solé, 2009), abarca entre sus objetivos el estudio de estructuras y fenómenos sociales entre los que se puede incluir, por supuesto, los procesos de

formación de la opinión pública. Como señala Marleti (Berrio, 2001: 98), en una sociedad de estas características, las corrientes de opinión tienen un componente de indeterminación que a menudo las hace aparecer como los fenómenos climáticos, es decir, sistemas complejos.

Probablemente, la principal razón del éxito o aplicación de las teorías de sistemas complejos a ámbitos, en principio alejados de sus orígenes como las Ciencias Sociales, es que los sistemas complejos comparten conceptos generales independientes del contexto en que se encuentren y tiene características y comportamientos universales que no varían a pesar de tratarse de sistemas muy diferentes (San Miguel, Toral y Eguíluz, 2004).

La característica principal de los sistemas complejos es que el comportamiento global del sistema no es reducible a la suma de sus partes o al comportamiento típico o promedio de los componentes del sistema. Si tomamos como ejemplo una sociedad, diríamos que su comportamiento global no es reducible, según la visión reduccionista clásica, a la psicología individual y las acciones de sus componentes, cuyas características individuales pueden llegar a ser irrelevantes para el comportamiento colectivo.

La segunda característica de los sistemas complejos es la que se relaciona con las transiciones de fase. El comportamiento complejo se refiere a una situación intermedia entre un estado bien ordenado y un estado totalmente desordenado. Esta se puede relacionar con los fenómenos de opinión que provocan la aparición repentina y momentánea de un tema en la agenda pública. Los populares “virales” que se producen en la Web 2.0, responden a una transición de fase en el sistema de comunicación que provoca la emergencia de un tema hasta entonces escondido entre la cantidad ingente de información que alberga la web.

El desarrollo en los últimos años de las investigaciones en Redes Complejas ha provocado el hallazgo de otros fenómenos relacionadas con el funcionamiento y características de estas. Por ejemplo, el viejo experimento de las cartas de Milgram, que había dado como resultado un nivel de separación entre todas las personas del planeta de seis grado, no se había teorizado de forma matemática hasta las investigaciones de Watts y Strogatz en 1998.

Estos dos investigadores comenzaron con un círculo de nodos, donde cada nodo está conectado a su inmediato vecino y al siguiente vecino más cercano. Para hacer este mundo más pequeño se modificaron algunos de los enlaces aleatoriamente. Estos enlaces de tipo largo ofrecieron los atajos cruciales entre grandes distancias dentro de la red, acortando drásticamente la separación entre los nodos que formaban la red.

Este tipo de redes, que denominaron redes de mundo pequeño, presentan la característica de que dos objetos pueden conectarse con pocos pasos intermedios. A medida que aumentamos el nivel de aleatoriedad en una red regular, el número de pasos para llegar de un extremo a otro de la red se reduce.

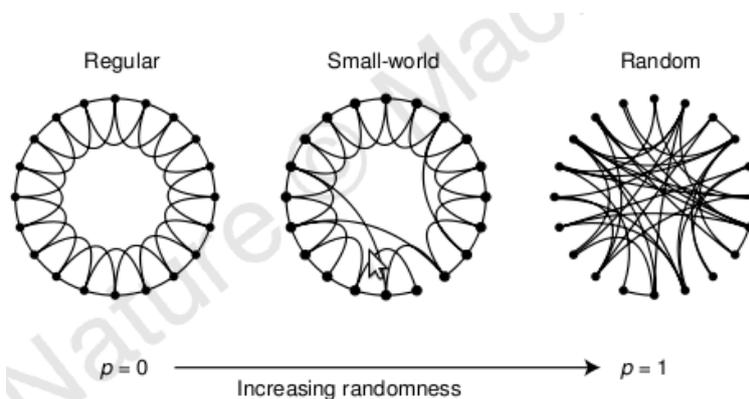


Figura 11: Red de mundo pequeño. Extraído de Watts y Strogatz (1998: 441).

Las investigaciones posteriores de Barabasi sobre el grado de separación e interconexión de las páginas existentes en Internet dieron como resultado un grado de separación que no alcanza los 19 pasos de promedio, lo cual demuestra que Internet es una red de mundo pequeño (Barabási, 2002: 34). Esta característica ha sido descubierta en redes de citación científica, en redes de colaboración de actores y en redes sociales en general por poner algunos ejemplos.

Otra de las características que presentan las redes complejas es que son redes libres de escala. En las redes que presentan esta propiedad, la probabilidad de que un nodo tenga K hiperenlaces es inversamente proporcional a la potencia K^γ , siendo γ un exponente característico de la red, lo que significa que hay un número pequeño de nodos que tiene un número de hiperenlaces muy alto. Este tipo de nodos que

presentan un alto índice de conexión es lo que se conoce en la terminología de Internet como *hubs*.

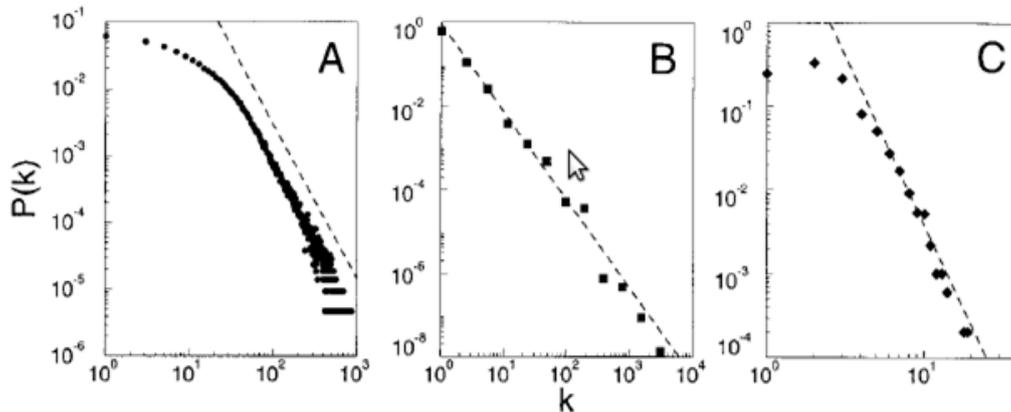


Figura 12: Funciones de distribución de conectividad en grandes redes. (A) Red de colaboración de actores $N = 212.250$ vértices en promedio de conectividad $(k) = 28.78$. (B) WWW. $N = 325.729$ $(k) = 5.46$ (6). (C) Red eléctrica. $N = 4941$, $(k) = 2.67$. Extraído de Barabási y Albert (1999: 510).

En este tipo de redes la distribución de los hiperenlaces sigue una curva de ley de potencia, lo que implica que los componentes que entran a formar parte de la red tienden a conectarse con los nodos más conectados. Según afirman Barabasi y Albert, este hecho se debe a dos mecanismos: (1) las redes se expanden continuamente por la suma de nuevos vértices, y (2) los nuevos vértices se conectan preferentemente a los sitios que están bien conectados. Un modelo basado en estos dos mecanismos reproduce esa distribución libre de escala, lo que indica que el desarrollo de grandes redes está gobernado por un fenómeno de autoorganización que va más allá de las particularidades de los sistemas individuales.

La que señalábamos como segunda característica de los sistemas complejos, es decir, las transiciones de fase que provocan el paso de un nivel de orden a otro diferente, se basan en dos fenómenos: (1) la autoorganización y (2) la emergencia. En las fases de transición, las que llevan al sistema de una forma de organización a otra, el sistema, de forma autoorganizada puede expandirse e intensificarse a sí mismo por un proceso de retroalimentación positiva, lo que provoca que pequeñas causas puedan generar transiciones de fase con grandes efectos en el sistema y su estructura.

La motivación principal para aplicar el concepto de sistema autoorganizado al ámbito de las ciencias sociales, y concretamente a los estudios de opinión pública, se basa en la hipótesis, defendida por muchos autores que citamos en este trabajo, de que el mundo moderno es inherentemente complejo y dinámico y sus fenómenos pueden ser mejor explicados con conceptos basados en las formas de organización en red.

Los sistemas sociales tienen un carácter dinámico que surge de la interacción entre actores individuales o grupos de actores y las estructuras sociales. Las sinergias que surgen de los procesos de comunicación entre actores humanos tienen como resultado la producción y reproducción de las estructuras sociales que, a su vez, permiten nuevas prácticas y comunicaciones por las que las estructuras sociales pueden volver a reproducirse. Este proceso es el que se conoce como autoipoesis social, según el cual tenemos dos niveles de autoorganización social:

- En un **nivel sincrónico** un sistema complejo se produce y reproduce por autoipoesis permanentemente.
- En un **nivel diacrónico** el orden emerge del desorden en puntos críticos de desarrollo del sistema, o como hemos definido anteriormente, en los puntos críticos de las fases de transición.

Los sistemas complejos se mantienen y producen permanentemente. Ahora bien, debido a contradicciones en el sistema o a posibles entradas de energía en caso de no tratarse de sistemas cerrados –como de hecho sucede en el sistema social– se producen fases de transición en las que aparecen nuevas cualidades del sistema. En el sistema comunicativo actual hay muchos más agentes que actúan como canales de comunicación, hay mucha más cantidad de información en circulación y ésta es más variada (López García, 2006: 231). Para el estudio de este tipo de sistemas es necesario un periodo de tiempo suficientemente extenso que permita no sólo fotografiar el sistema en un momento dado, sino estudiar la dinámica del mismo para reconocer patrones de comportamiento que nos permitan establecer los puntos críticos en las fases de transición y los estados de equilibrio.

Las transiciones de fase entre un estado de equilibrio del sistema y otro, dependen de esos procesos de retroalimentación y reverberación del sistema. En un sistema complejo se pueden producir dos tipos de retroalimentación, negativa y positiva.

- La **retroalimentación negativa** supone una autorregulación del sistema. A través de diversos mecanismos, un sistema complejo puede detectar las variables externas y autorregularse para llegar a un punto de equilibrio en el que las condiciones son adecuadas para el sistema⁷. Este tipo de sistema con procesos de retroalimentación negativa es lo que se conoce como sistemas adaptativos. Ante una variable exterior se produce un fenómeno emergente que adapta el sistema a las nuevas condiciones.
- La **retroalimentación positiva** en un sistema se produce cuando una variable externa provoca una reacción en el sistema que se retroalimenta y para generar una reverberación⁸. Este tipo de sistemas sin mecanismos de retroalimentación negativa o autorregulación son los que se conocen como sistemas no adaptativos. Ante la aparición o modificación de una variable exterior se produce un fenómeno emergente en el sistema, que no alcanza un nuevo punto de equilibrio sino que produce una reverberación que solo se detendrá cuando desaparezca la variable que lo produjo, o bien, cuando aparezca otra variable que lo contrarreste.

Los trabajos de análisis de redes sociales en la web han utilizado conceptos como esfera web o red temática asociados a la organización temática de la información. Sin embargo, la simple organización en temas no aporta toda la información necesaria para conocer el sistema de comunicación en la web. Aunque puede decirse que esta red de conexiones entre sitios web muestra la estructura del sistema comunicativo, es necesario conocer igualmente otros factores que determinan los procesos de selección de la información.

⁷ Un termostato es el ejemplo más simple de retroalimentación negativa. Cuando la temperatura sube, el termostato lo detecta y regula la fuente de calor para conservar la temperatura de equilibrio en el sistema.

⁸ El ejemplo más simple de este tipo de retroalimentación positiva es el de la guitarra eléctrica que al ser acercada al altavoz por el que sale su sonido, capta dicho sonido creando un bucle que provoca una reverberación o acople acompañado de un sonido ensordecedor.

Existen procesos de selección de la información propios de Internet entre lo que podemos considerar a la visibilidad como el principal factor en la determinación de la capacidad de influencia de los actores que participan en una red. Pero al mismo tiempo, los factores que influyen en el nivel de visibilidad pública en Internet son complejos y necesitan un gran conocimiento del sistema y recursos para modificarlo a nuestro favor.

La visibilidad tiene diferentes definiciones dependiendo del ámbito desde el que se estudia. Así, en el ámbito del marketing se entiende por visibilidad la presencia de una marca o producto en el medio del consumidor (Dreze y Zufryden, 2003: 25), mientras que en el ámbito del análisis de redes sociales, se entiende por visibilidad el potencial para comunicar opiniones (Freeman, 1978: 219-220) y con ello influir en otros actores.

Dreze y Zufryden, desde la perspectiva del marketing, distinguen dos formas de visibilidad para una empresa u organización: la visibilidad *online* y la visibilidad *offline*. Cada una de estas formas de visibilidad está influida por de una serie de factores

- La visibilidad online depende de factores como la publicidad, el posicionamiento en buscadores, las referencias a través de hiperenlaces o menciones en otras páginas web, los resúmenes de noticias o los correos electrónicos.
- La visibilidad offline, en cambio, depende únicamente de la publicidad y las noticias.

Para realizar un cálculo de la visibilidad, los autores desarrollan una serie de representaciones matemáticas que permiten otorgar un valor numérico a cada uno de estos factores y relacionarlos después en una fórmula absoluta que representa la visibilidad total de un actor.

Desde esta misma perspectiva de la gestión empresarial y el *marketing*, Schmitz-Manz y Gaul (2005) consideran que la visibilidad está compuesta de diferentes partes. Estas partes son la visibilidad obtenida a través de hiperenlaces que vienen de otras páginas, la visibilidad conseguida a través de la aparición en directorios online y la visibilidad proveniente de los motores de búsqueda. Todas estas formas de visibilidad

y cualquier otra que pueda surgir, se pueden clasificar, según los autores, en dos categorías de factores que la determinan.

- Los factores psicológicos de la visibilidad *online*. Relacionados con la forma en que las personas buscan, navegan e interactúan con Internet. Por ejemplo, el hecho de que sólo se consultan los diez primeros resultados de un buscador tiene que ser tomado en cuenta.
- Los factores físicos de visibilidad *online*. Estos son los enlaces que recibe una página de otras páginas web, los anuncios publicitarios que contrata o el posicionamiento en buscadores.

Sean cuales sean los factores que intervienen en la visibilidad de un actor, parece claro que los actores con más recursos pueden obtener un mayor nivel de visibilidad en la esfera pública. La mayoría de los actores no están en posición de afectar el proceso político de forma directa sino que deben hacerse visibles a través de otros actores más importantes consiguiendo el apoyo de estos. Solo a través de estas alianzas podrán ejercer presión en las tomas de decisión políticas.

El sistema comunicativo actual dominado por la web y la tecnología digital ofrece a los actores un nuevo potencial para evitar a los medios de comunicación de masas tradicionales y conseguir de manera directa cierta visibilidad pública a través de su presencia en Internet. Con una inversión mínima de recursos, cualquiera puede crear una página web y hacer públicas sus opiniones para todo el mundo. De esta forma, y en línea con la idea de comunicación masiva individual que planteaba Castells, usuarios individuales o sitios web con ideas y puntos de vista similares pueden crear ciertos flujos de información y opiniones en Internet mediante la creación colectiva de enlaces a otros sitios web (Koopmans y Zimmerman, 2003: 5-6).

Desde la perspectiva de la comunicación política, Koopmans y Zimmerman (2003) defienden que los principios sobre los que se asienta el establecimiento de la visibilidad pública en Internet operan mediante los dos mecanismos principales de búsqueda de información: (1) selección vertical, en la que participan portales y buscadores, y (2) selección horizontal, determinada por las redes de hiperenlaces.

En el caso de los mecanismos de selección vertical, los portales y buscadores guían al usuario presentando una selección jerárquica de sitios web que trata de ser relevante para el tema que el usuario ha definido mediante los criterios de búsqueda introducidos en el buscador. De esta manera, se puede decir que los buscadores actúan como *gatekeepers* al ofrecer una selección de la realidad de acuerdo con unos criterios propios.

En el caso de los mecanismos de selección horizontal, una red de selección formada por hiperenlaces ayuda al usuario a encontrar sitios web siguiendo los enlaces que se ofrecen en otros sitios web. A través de hiperenlaces los actores públicos actúan ellos mismo como *gatekeepers* en la esfera online al ofrecer enlaces a unos actores determinados y excluir a otros de su selección. (Koopmans y Zimmermann, 2007: 219-221). Esta idea de selección horizontal cobra sin duda más relevancia desde la popularización de plataformas sociales de comunicación como Facebook y Twitter que en los últimos años se han convertido en muchas ocasiones en redes personalizables de selección y acceso a todo tipo de información, incluidas entre ellas las noticias.

En los procesos de selección de la información que hemos denominado horizontal, el hiperenlace supone la principal tecnología que determina este tipo selección. Las redes de hiperenlaces comparten características con los fenómenos estudiados por el análisis de redes sociales desde la década de los 60 y han sido estudiadas aplicando algunos de los conceptos creados por esta disciplina.

En el ámbito del análisis de redes sociales, Freeman, en su artículo “La centralidad en las redes sociales”, define el nivel de centralidad de un actor en la red como factor fundamental que determina su capacidad de influencia, su visibilidad o su potencial para comunicar sus opiniones. De alguna manera es un foco de comunicación, al menos respecto a aquellos con los que está en contacto, y está en posición de considerarse parte del canal principal del flujo de información dentro de la red (Freeman, 1978: 219-220).

La idea de centralidad aplicada a la comunicación humana fue introducida por Bavelas en 1948 y según esta idea, una persona que está en una posición que le permite tener contacto directo con muchas otras debería comenzar a verse y a ser vista por otros como un canal principal de información. De esta forma, las conexiones entre actores

pueden reflejar corrientes cognitivas, culturales y sociales y formaciones que incluyen la emergencia de redes y la difusión de ideas sobre ciertos temas (Bjorneborn en Lee, 2006: 51).

Si bien el número de enlaces que recibe un sitio web constituye un indicador básico de la visibilidad, el problema con esta medida es que no tiene en cuenta quién realiza el enlace. Una medida más sofisticada de la visibilidad es la que aporta el algoritmo HITS de Kleinberg (1999) que se basa en la premisa de la existencia de dos tipos distintos, pero relacionados, de sitios web: los sitios de autoridades y los *hubs*.

- Los sitios de autoridades son sitios muy referenciados, es decir con muchos enlaces entrantes relacionados con un tema.
- Los *hubs*, en cambio, son páginas que apuntan a las páginas de autoridades y con ello les otorgan autoridad.

Existe una relación de refuerzo entre los dos tipos de sitios. Un *hub* prominente apuntará a muchos sitios de autoridades y un sitio de una autoridad importante será apuntado por muchos *hubs*.

Cuando entendemos los enlaces hiperenlaces como un tipo de relación, podemos utilizar los métodos de análisis de redes sociales para describir los patrones de conexiones o desconexiones observados en una red de sitios web (Park, 2003:48; Jackson, 1997:292). Desde los años 60, los sociólogos han estudiado las propiedades de las redes sociales y los conceptos propuestos por ellos están siendo utilizados con algunas adaptaciones al nuevo sistema comunicativo que supone Internet.

Uno de los conceptos desarrollados por la teoría de redes sociales que se aplica al estudio de redes sociales online es el de lazos fuertes y lazos débiles que desarrollara Granovetter en 1973. Según esto una red social puede estar dividida en lazos débiles y fuertes, siendo estos últimos los que presentan un nivel de concentración mayor.

El concepto de centralidad es otro de los conceptos aplicados al estudio de redes sociales online. La medición de la centralidad puede indicar el grado de concentración de hiperenlaces en determinados sitios web o determinados grupos de sitios web

dentro de una comunidad más amplia. El análisis hiperenlace puede aportar también una medida de la posición de influencia de un sitio web, identificando sitios que aportan conexiones cruciales entre grupos de sitios web, incluso en casos en que estos sitios web conectores no tienen una gran cantidad de enlaces apuntando hacia ellos. Buscadores como Google utilizan la medición de la posición estructural de un sitio web a través de la estructura *hiperenlace* que rodea al sitio web como criterio para determinar cuál es el sitio web más influyente dentro de un grupo de sitios web (Henzinger, 2001:384). De esta manera se ha utilizado la medición del número de enlaces como medida de la visibilidad o centralidad de un sitio o un grupo de sitios web.

Otro de los conceptos adaptados al estudio de redes online es el de propagación. La propagación hace referencia al alcance y velocidad de difusión de la información entre redes de comunicación. O dicho en otras palabras, la medida de la capacidad de contagio que tiene una información (Bonchek, 1997:33). La fuerza de relación hace referencia a la densidad de la conexión entre diferentes redes temáticas o esferas web. Esta densidad, también llamada conectividad, influye en la propagación de un tema u opinión entre dos esferas web o dentro de una propia esfera (Jackson, 1997).

En estas consideraciones preliminares en torno a las redes temáticas como campo y objeto de estudio hemos visto diferentes aproximaciones al análisis de las relaciones sociales, de las redes de comunicación y de la formación de los fenómenos de opinión. Hemos presentado conceptos como redes de hiperenlaces, visibilidad o centralidad y hemos propuestos el concepto de red temática como la alternativa que consideramos más adecuada para llevar a cabo estudios de opinión pública en la web teniendo en cuenta las características del sistema comunicativo actual. A continuación, en la Tabla 7, presentamos un resumen comparado de estas diferentes tendencias y conceptos.

Ámbito de investigación	Análisis de redes sociales	Análisis hiperenlaces	Análisis de redes temáticas
Actores/nodos	Individuos/contactos	Sitios web/ <i>hiperenlaces</i>	Sitios web/ <i>hiperenlaces</i>
Objetivos de investigación	Relaciones sociales, posición social entre actores	Patrones de relación/posición estructural de sitios web	Naturaleza de las relaciones/Formación de subgrupos en una red
Comunicación/Nivel de análisis	Individual/ contactos individuales	Individual/relaciones individuales	Colectivo/ Recopilación de enlaces (subred/esfera)
Canal de comunicación	Comunicación Interpersonal	Comunicación Interpersonal	Comunicación de masas
Conceptos principales	Centralidad/ densidad	Centralidad/ densidad	Densidad/ tamaño

Tabla 7: Resumen comparado de las teorías de análisis de redes sociales, de análisis hiperenlace y de análisis de redes temáticas. Adaptado de Lee (2006:71).

Proponemos la utilización del concepto de red temática y su aplicación en los estudios de opinión pública en la web pues entendemos que es una forma adecuada de explicar los nuevos fenómenos de comunicación que posibilita el sistema comunicativo actual pero ¿Qué son exactamente las redes temáticas? ¿Qué componentes las conforman? ¿Cuáles son las características de esos componentes? En el siguiente apartado tratamos de definir claramente el concepto de redes temáticas para el que realizamos una propuesta de componentes y caracterización que introducimos con una revisión previa de las principales fases de desarrollo del concepto de redes temáticas aplicado a los estudios en la web.

2.2. Redes temáticas en la web: una propuesta de caracterización y componentes

En el apartado anterior analizábamos los orígenes del concepto de redes temáticas en el ámbito de la Comunicación, en el de las Ciencias Políticas y en el de la Sociología. Sin embargo, el antecedente más directo de la aplicación del concepto de red temática en el análisis de la opinión pública en la web lo encontramos en el análisis de redes de hiperenlaces que, como veremos a continuación, se ha aplicado en muchos y variados ámbitos científicos.

El análisis de hiperenlaces se ha empleado en el ámbito de la física (Barabasi, 2002), la informática, (Brin y Page, 1998) o las ciencias de la información (Almind y Ingwersen, 1997). En el ámbito de la comunicación social se ha aplicado a investigaciones sobre la comunicación entre representantes políticos (Ackland, 2005; Park y Thelwall, 2008), los grupos de interés (Rogers y Ben-David, 2008; Shumate y Lipp, 2008; Shumate, 2012), o los flujos de información entre países (Hsu y Park, 2012).

Sin embargo, a día de hoy parece no existir todavía un consenso unificado sobre cómo interpretar los resultados que ofrecen los análisis de hiperenlaces. La naturaleza dinámica de la web y la falta de controles de publicación hace muy difícil extraer conclusiones claras de los datos cuantitativos que este tipo de análisis pueda aportar (Thelwal, 2006).

La dificultad de obtener muestras de información suficientemente grandes, la volatilidad de las muestras extraídas provocada por la constante publicación de información nueva, o fenómenos como el denominado *rich get richer*, por el cual los sitios más enlazados reciben cada vez más enlaces entrantes, sin atender a las razones lógicas que se suelen atribuir a la creación de hiperenlaces, hacen muy difícil desarrollar una teoría unitaria del análisis de hiperenlaces y se requieren por tanto metodologías combinadas que puedan resolver estos problemas de investigación.

Precisamente esta es la problemática a la que pretende hacer frente la aplicación del concepto de red temática en el estudio de los procesos de formación de la opinión pública en la web. Para ello, proponemos utilizar un modelo dual que diferencia y

aplica a la vez las nociones de redes *empíricas* y de redes *epistemológicas*. De este modo, se relaciona el análisis de redes de hiperenlaces empíricas, es decir, aquel que tiene en cuenta únicamente los hiperenlaces (basadas en conexiones en el código fuente entre diversas páginas web) con el análisis de redes temáticas epistemológicas (basadas en las relaciones que se dan entre los actores que participan del debate) es decir, aquel que analiza las relaciones que construyen diferentes actores en torno a un tema de interés común (Young y Leonardi, 2012).

En el ámbito de la sociología, la rama que estudia las redes se conoce como Análisis de Redes Sociales, (Wasserman y Faust, 1994) y desde este ámbito se han propuesto una serie de conceptos que están siendo utilizados con algunas adaptaciones en el análisis del nuevo sistema comunicativo que supone Internet. En el Capítulo 3 “Sistema de análisis de redes temáticas en la web” analizaremos con detalle las medidas y algoritmos más utilizados en el análisis de redes sociales y cuáles de ellos pueden ser considerados los más adecuados para estudiar las redes temáticas en la web. A continuación elaboramos una pequeña introducción a este tipo de medidas pues lo consideramos imprescindible para comprender el funcionamiento de los sistemas de análisis de información en la web y en concreto para comprender los sistemas de análisis de la opinión pública incluido el análisis de redes temáticas en la web.

El análisis de hiperenlaces puede aportar una medida de la posición de influencia de un sitio web, identificando sitios que aportan conexiones cruciales entre grupos de sitios web, incluso en casos en que estos sitios web conectores no tienen una gran cantidad de enlaces apuntando hacia ellos. Buscadores como Google utilizan o han utilizado la medición de la posición de un sitio web a través de la estructura de hiperenlaces que le rodea como criterio para determinar su capacidad de influencia en la red (Henzinger, 2001: 384).

Cuando entendemos los hiperenlaces simplemente como una forma de relación, podemos utilizar los métodos de análisis de redes sociales para describir los patrones de conexiones o desconexiones observados en una red de sitios web (Park, 2003: 48; Jackson, 1997: 292). Centralidad, densidad o intermediación son solo algunos ejemplos de los parámetros de análisis y medición que utiliza el análisis de redes

sociales. Como ya dijimos anteriormente estudiaremos con mayor profundidad estas medidas de análisis en el Capítulo 4.

Si a esta aplicación de los conceptos del Análisis de Redes Sociales en el análisis de hiperenlaces en la web añadimos la categorización temática que proponía Luhmann, nos encontramos con conceptos como esfera web (Rogers, 2002: 198) o red temática (Foot y Schneider, 2002: 226).

El concepto de red temática desarrollado por Rogers (2002) se define como una red de hiperenlaces por la que fluye información relacionada con una palabra clave. Como vemos, plantea un estudio del sistema comunicativo en Internet centrado en los circuitos de comunicación y en los flujos de información. Propone además diferenciar los circuitos en función de palabras clave y organizados, por tanto, en torno a temas.

El concepto de esfera web, a diferencia del de red temática, se plantea no sólo como una colección de sitios web que tratan un tema en común sino como una serie de recursos digitales que se organizan dinámicamente en torno a un evento, concepto o tema, y a menudo están conectados a través de hiperenlaces (Foot y Schneider 2002: 227).

Los conceptos de red temática y esfera web se han utilizado en numerosos trabajos con un significado común o al menos muy cercano. En algunos casos se ha considerado el concepto de esfera web como una categoría inferior a la red temática, delimitada por un mayor nivel de relación dentro de una red temática más amplia (Lee, 2006). En otros casos el concepto de esfera web se ha utilizado como categoría amplia dentro de la cual existen unidades más pequeñas organizadas en torno a temas, subtemas o marcos de referencia que tienen como resultado un mayor nivel de conectividad y relación (Foot y Schneider, 2002).

En la práctica de las investigaciones el concepto de esfera web es cuasisinónimo del de red temática. En este estudio se utilizará el término red temática por ser éste más cercano a la teoría de redes en la que se basa la argumentación de la propuesta metodológica. En cualquier caso, ambos conceptos comparten el planteamiento de una organización en categorías y subcategorías que se delimitan por el hecho de

compartir interés por un tema u opinión común, lo cual se refleja en un mayor nivel de conectividad entre nodos.

Utilizando el concepto de red temática, autores como Ackland, Thelwall o Park han desarrollado estudios con el objetivo de analizar diferentes regiones de Internet como la blogosfera de los partidos demócrata y republicano en Estados Unidos (Ackland, 2005), la visibilidad de la nanotecnología en Internet (Ackland, 2007), o los hiperenlaces que reciben y realizan las páginas de los parlamentarios en Corea del Sur (Park y Thelwall, 2008).

Esta forma de estudio de la opinión pública resulta especialmente interesante para nuestro trabajo, ya que si utilizamos esta perspectiva para visualizar el sistema comunicativo en Internet, veremos infinidad de esferas o redes temáticas que se dividen a su vez en subesferas o subredes, que en ocasiones se encuentran conectadas entre sí formando una red de redes temáticas que forman una macroestructura temática.

Esta idea nos hace volver al concepto de opinión pública que planteaba Luhmann, aunque esta vez con una estructura mucho más fragmentada tanto en relación a la cantidad de redes temáticas existentes –opinión pública fragmentada o grupos de opinión– como a la cantidad de emisores de información que pueden influir en el proceso de formación de la opinión pública.

El principal objetivo del análisis de redes temáticas es aportar información sobre cómo funcionan los procesos de formación de opinión en un sistema de estas características, así como extraer datos objetivos de las corrientes de opinión que se generan en torno a un tema a lo largo del tiempo.

Para conseguir que el análisis de redes temáticas en la web pueda ser aplicado con la agilidad suficiente y que así se adapte a la dinámica característica de los fenómenos de opinión en la web, es necesario establecer metodologías claras y sistemáticas, que puedan ser: (a) escalables y (b) reproducibles con una periodicidad de tiempo razonablemente corta.

El problema, tal como lo entendemos nosotros, es que en los estudios que utilizan el análisis de hiperenlaces, los conceptos de tema y actor no se han definido claramente a pesar de haber sido utilizados y considerados de forma implícita como elementos importantes en el análisis de redes temáticas.

A su vez, los temas en torno a los que se forman las redes de hiperenlaces y los actores que, en última instancia, son los responsables de crear esas redes de hiperenlaces y de aportar las opiniones relacionadas con los temas, son factores fundamentales en la constitución de redes temáticas en la web.

Consideramos que únicamente tras definir claramente cuáles son los elementos que caracterizan a una red temática podremos proponer sistemas de análisis de las mismas. Por todo ello y para tratar de superar el problema de indefinición que hemos señalado, dedicamos los próximos apartados a estudiar las características de los tres componentes que intervienen en nuestra propuesta teórica de modelo de red temática: (1) redes de hiperenlaces, (2) temas y (3) actores. Al final de cada apartado se hará una propuesta de definición propia para cada uno de los tres conceptos señalados.

2.2.1. Redes de hiperenlaces

A mediados de los 90, Almind e Ingwersen (1997) acuñan el término *webometrics* para referirse a un conjunto de técnicas de investigación bibliométrica que aplican a los estudios de la web, entre ellos el estudio de hiperenlaces. Recordemos que entendemos por hiperenlace la porción del código fuente de una página web (elemento `<a>` más atributo `href`) que activa un nuevo destino de navegación si el usuario hace clic sobre el mismo.

Desde la perspectiva de las Ciencias de la Información, un hiperenlace actúa de forma similar a una cita bibliográfica, ya que implica que el autor que realiza el enlace atribuye cierta relevancia a la página de destino. De esta forma, se pueden explicar las redes creadas a través de hiperenlaces en términos de credibilidad, prestigio y confianza (Kleinberg, 1999) de la misma manera que se hace con las citas bibliográficas en los estudios de cienciometría. Otra forma de ver los hiperenlaces, en la terminología de los fundadores del buscador Google, es en forma de voto: si la página A enlaza a la página B, la página A está emitiendo un voto por la página B. De

este modo, siempre según Google, las páginas con muchos enlaces de entrada se consideran páginas de calidad o de autoridad en un tema.

Ahora bien, más allá de la visión inicial que equiparaba el hiperenlace con la cita bibliográfica, la teoría de redes sociales señaló otras funcionalidades del hiperenlace asumiendo que el patrón de relaciones que se establece en la web tiene influencia sobre el comportamiento de los actores que establecen un hiperenlace (Freeman, 2004).

De la aplicación de estos conceptos provenientes de la teoría de redes sociales surge una nueva perspectiva en el estudio del hiperenlace: el análisis de redes de hiperenlaces (Park, 2003). Este tipo de análisis se ha aplicado al estudio de regiones de Internet tales como la esfera política de los Estados Unidos (Foot y Schneider, 2002), la visibilidad de la nanotecnología en Internet (Ackland et. al., 2007), o los sitios web de los políticos en Corea del Sur (Park y Thelwall, 2008).

Según este enfoque surgido de los estudios de comunicación, en una red de hiperenlaces los actores son las organizaciones o los individuos que editan o administran los sitios web y los hiperenlaces son un mecanismo de asociación entre creadores de información. En este tipo de redes, los hiperenlaces no son meros mecanismos para otorgar credibilidad a la manera de las citas bibliográficas, sino que, tal y como señalan Ackland y Gibson (2006), tienen al menos cinco funciones adicionales muy importantes:

- **Provisión de información:** los hiperenlaces pueden dirigir a los visitantes hacia fuentes de información adicionales.
- **Construcción de redes y refuerzo:** los hiperenlaces permiten a las organizaciones crear alianzas virtuales y redes y mantener y reforzar las relaciones existentes en el ámbito offline a través de los lazos creados online.
- **Construcción de identidad:** la creación de conexiones indirectas o implícitas entre grupos puede ayudar a fortalecer la identidad de los componentes de esos grupos a través del fortalecimiento de la identidad del propio grupo.

- **Intercambio de audiencia:** la integración en un grupo puede conseguir de manera más eficiente e inmediata que la audiencia y los potenciales apoyos se compartan entre los integrantes del grupo especialmente entre los grupos de opinión.
- **Amplificación del mensaje** o fuerza de multiplicación: a través de los hiperenlaces se transmite a los usuarios una imagen distorsionada de la extensión del apoyo con que cuenta el mensaje.

Por tanto, los hiperenlaces además de servir al flujo de información y la transmisión de mensajes, tienen una función representativa que influye en los procesos identitarios de los actores que los realizan y permiten construir lazos de comunidad (Shumate, 2012).

Definición 1: Redes de Hiperenlace

Desde una perspectiva combinada de las ciencias sociales y las ciencias de la información, una red de hiperenlaces es el sistema dinámico que hace posible el flujo de información entre diferentes sitios web y a través del cual se establecen relaciones de comunidad entre los actores que los crean.

2.2.2. Temas

Al aplicar un enfoque comunicacional de este tipo, surge la necesidad de incluir en el análisis otros factores más allá del propio hiperenlace; analizar los actores que intervienen en la red, ya sean estos individuos u organizaciones, y analizar el tema en torno al que se asocian.

Los primeros intentos de definir el concepto de tema en el ámbito de la comunicación vienen de la teoría de la *agenda setting*. McCombs y Shaw, en su célebre artículo "The *Agenda setting* Function of Mass Media", definían la función de la *agenda setting* como el resultado de la relación que se establece entre el énfasis manifestado por el tratamiento de un tema por parte de los *mass media* y las prioridades temáticas manifestadas por los miembros de una audiencia tras recibir los impactos de los medios de comunicación (McCombs y Shaw, 1972). El mismo año, Cobb y Elder

definían la idea del tema como conflicto o motivo de desacuerdo entre dos partes (Cobb y Elder, 1972).

En el ámbito de la sociología, Luhman (1974) desarrolla el concepto de tema, según el cual, la opinión pública debe ser concebida como estructura temática de la comunicación pública. Para Luhman, de la transformación estructural del sistema político se ha derivado la creciente centralidad de los temas y de las controversias políticas (*issues*), y por consiguiente, los temas se han convertido en recursos estratégicos esenciales de la política en las sociedades complejas.

Dentro de la misma corriente sociológica, Marletti (1985) diferencia entre tema y tema político o *issue*. Según este planteamiento, compartido también por Shaw y otros, el tema no es sino una forma de generalización simbólica de situaciones singulares que posibilita la comunicación, mientras que un *issue* o tema político se forma a través de la consolidación de las posturas defendidas en una controversia generada en torno a un tema. Por tanto, los temas políticos o *issues* son objetos de discusión que se desarrollan a través de la interacción entre actores sociales y políticos con distintas posiciones.

El término de red temática con el que trabajamos en este artículo es una traducción del término en inglés *issue network* que hace referencia al concepto de *issue* que acabamos de definir. No existe en castellano la diferenciación que existe en inglés entre *topic* e *issue*, y de ahí la necesidad de definir con precisión el concepto al que nos referimos cuando hablamos de tema.

El objeto de estudio en el análisis de redes temáticas son estos temas de controversia o políticos (*issues*), y el análisis de redes de hiperenlaces es la técnica que nos permite determinar las relaciones empíricas existentes entre los actores que defienden sus posiciones en torno a un tema.

La navegación entre hiperenlaces puede ayudar a comprender las relaciones que existen entre diferentes actores y entre las distintas posiciones respecto a un tema en discusión y aportar un contexto más amplio de las diferentes voces que participan en un espacio temático (Rogers y Marres, 2000; Shumate y Dewitt, 2008; Young y Leonardi 2012). Es decir, a través de la navegación entre hiperenlaces podemos

acceder a las diferentes opiniones y discursos de los actores que participan en la red y obtener una imagen global del tema en discusión, obteniendo además información sobre los actores que hay detrás de los diferentes discursos.

Es de este modo que el análisis empírico de las redes de hiperenlaces, que estudia los flujos de información, puede dar paso a un análisis epistemológico que busca desvelar el valor representativo de los hiperenlaces y de la red que forman. Los temas, entendidos como objetos de discusión que afectan a la sociedad y que se desarrollan a través de la interacción entre actores con distintas posiciones, se convierten así en objeto de estudio identificable.

Definición 2: Temas

Un tema es un conjunto de ideas, teorías o conceptos que generan controversia entre grupos de actores políticos, sociales o económicos.

2.2.3. Actores

En la definición precedente de tema señalamos como necesaria la interacción entre actores con distintas posiciones, pero ¿quién o qué debe ser considerado un actor? O en otras palabras: ¿Bajo qué circunstancias un individuo, una organización o incluso una red deben ser considerados un actor?

Solo entendiendo la naturaleza de los actores que participan en la red y cómo generan los diferentes discursos y defienden sus posiciones, podemos entender el funcionamiento de las redes temáticas en la web.

El concepto de actor que proponemos surge de la literatura de los movimientos sociales (Jenkins, 1995), los grupos de interés (Berry, 1984) y de la literatura de los *stakeholders* o partes interesadas (Freeman, 1984). Consideramos que en los tres ámbitos el fenómeno que se estudia presenta un gran nivel de similitud y es de alguna manera intercambiable: aquellos que defienden una determinada demanda. Utilizamos el término actor en un intento de abarcar estos tres ámbitos utilizando un único término

que pueda incluir (1) tanto a los individuos u organizaciones que forman parte de un movimiento social, como (2) un grupo de interés o (3) partes interesadas que actúan como *stakeholders* frente a una empresa privada.

Defendemos, por tanto, que la teoría de redes temáticas que presentamos en este artículo puede ser aplicada, siempre con las adaptaciones que convengan en cada caso, en investigaciones relativas tanto al ámbito de la sociología, las ciencias políticas como las ciencias empresariales. Consideramos que resultaría confuso utilizar un término diferente para identificar a los individuos u organizaciones que intervienen en las diferentes redes temáticas y por ello proponemos un término que sea aplicable en los tres ámbitos de estudio.

Tanto los estudios de los movimientos sociales (Sociología) como los estudios de los grupos de interés (Ciencias Políticas) consideran a aquellos que plantean reclamaciones dentro del ámbito de lo público, es decir, poniendo en relación directa a la ciudadanía con el estado (Berry, 1984; Jenkins, 1995). En estos casos, el interés recae en los grupos u organizaciones que tratan de influir en el poder legislativo para desarrollar leyes que defiendan sus intereses. Sin embargo, las ciencias empresariales en general, y los estudios de *stakeholders* en particular, consideran a aquellos que plantean reclamaciones dentro del ámbito de lo privado, es decir, aquellos grupos que tratan de influir en los procesos de decisión de las corporaciones (Freeman, 1984).

Como vemos, los factores que se tienen en cuenta a la hora de considerar a un individuo u organización como actor de una red temática varían en función de si este individuo u organización actúan en el ámbito de lo público o lo privado. De la misma forma, los factores que se tienen en cuenta varían cuando se estudia el concepto de actor desde diferentes perspectivas teóricas o de análisis.

Según Frooman (2010), son tres las teorías que analizan el concepto de actor: la teoría identitaria, la teoría económica y la teoría política. Según la teoría identitaria, todo lo que necesita un individuo u organización para ser considerada parte de una red temática es una queja o demanda relacionada con el tema de discusión; según la teoría económica, un actor puede ser considerado parte de una red temática siempre que tenga recursos para participar en ella; y finalmente, según la teoría política, un actor formará parte de una red temática si tiene una oportunidad de participar en ella.

Como vemos, la identificación de los actores atiende a diferentes factores en función del ámbito en el que participa el actor, público o privado; y en función del ámbito de estudio desde el que se realiza el análisis, teoría identitaria, teoría económica o teoría política. Sin embargo, y en un ejercicio de simplificación, proponemos una definición inclusiva que permita abarcar estas diferentes perspectivas y ámbitos de estudio y ofrezca una herramienta flexible aplicable a diferentes casos y ámbitos de aplicación.

Definición 3: Actor

Un actor es todo individuo u organización, del ámbito público o privado, que tenga quejas, recursos u oportunidad de influir en un tema de discusión.

2.2.4. Redes temáticas

En el presente capítulo se ha señalado la utilidad de aplicar al análisis de opinión en la web una teoría de redes temáticas con una perspectiva dual que tenga en cuenta no solo el análisis empírico de los flujos de información que se producen en las redes de hiperenlaces, sino también un análisis epistemológico que desvele el valor representativo de la interacción entre actores con distintas posiciones respecto a un tema de discusión.

Se ha propuesto una clarificación de los conceptos red de hiperenlaces, tema y actor, que son utilizados habitualmente en los estudios de análisis de opinión en la web y que por tanto podrá ser de utilidad a cualquier investigador de este ámbito con independencia de que utilice o no la teoría de redes temáticas que proponemos en su conjunto.

La caracterización de estos tres conceptos nos permite proponer como parte de nuestras conclusiones una definición operativa de red temática que presentamos en el cuadro de definición siguiente:

Definición 4: Red Temática

En el ámbito de la web, y siempre desde la perspectiva combinada de las ciencias sociales y las ciencias de la información, proponemos que una red temática es un conjunto de individuos u organizaciones que relacionados a través de una red de hiperenlaces plantean opiniones y utilizan recursos u oportunidades para influir en un tema de discusión.

3. HERRAMIENTAS PARA EL ANÁLISIS DE REDES TEMÁTICAS EN LA WEB

El objetivo de este capítulo consiste en **Identificar, describir y evaluar las herramientas para el análisis de redes temáticas en la Web**. Para ello dividimos este capítulo en tres secciones.

En el apartado **3.1. Consideraciones previas entorno a las herramientas para el análisis de redes temáticas en la web** estudiamos el origen de este tipo de herramientas y analizamos su relación con diferentes ámbitos de conocimiento como la Ingeniería Informática, las Ciencias de la Información o la Comunicación Social; todo ello con el objetivo construir un breve marco teórico que sirva de contexto al uso y desarrollo de este tipo de herramientas.

En el apartado **3.2. Caracterización y evaluación de herramientas para el análisis de redes temáticas en la web** nos proponemos describir y caracterizar los principales crawlers destinados a la investigación científica desde la perspectiva de un usuario investigador y para ello dividimos este apartado en la sección **3.2.1. Perspectivas y modelos de calidad en la evaluación de software**, en el que estudiamos las principales perspectivas en la evaluación de la calidad del software, las tendencias actuales en este ámbito y los principales modelos de calidad para la evaluación de software y en la sección **3.2.2. Selección y caracterización de herramientas**, en la que describimos las características y funcionalidades de las herramientas seleccionadas.

Finalmente en el apartado **3.3. Propuesta de modelo de calidad y evaluación de herramientas**, proponemos una herramienta de evaluación con una selección de parámetros e indicadores que nos permitan obtener una evaluación cualitativa de las herramientas seleccionadas.

El sistema de análisis de redes temáticas que proponemos, y que presentaremos en detalle a continuación se estructura en tres niveles: (1) análisis de redes de hiperenlaces, (2) análisis de opinión de los contenidos extraídos de esas redes (3) análisis de los actores que participan en la red. En las tres fases, las herramientas utilizadas para lograr los objetivos que nos planteamos serán web crawlers, también

conocidos como robots o *spiders*. Por esta razón, nuestro estudio se limita a estudiar los web crawlers y no incorpora otros tipos de herramientas que podrían estar relacionadas con el análisis de redes temáticas en la web como visualizadores gráficos de redes, analizadores automáticos de textos o herramientas de flujos de trabajo. Los crawlers actuales incorporan complementos que permiten desarrollar esas funciones y por tanto dejamos un estudio más amplio de otras herramientas para futuras investigaciones.

3.1. Consideraciones previas entorno a las herramientas para el análisis de redes temáticas en la web

Recordemos que la World Wide Web (W3 o WWW) es un sistema de recuperación de información con el propósito de dar acceso a un gran número de documentos hipermedia. Desde el punto de vista tecnológico, la Web se construye a partir de un modelo cliente-servidor en el que los usuarios utilizan un programa (cliente) para conectar con un ordenador remoto (servidor) en el que se almacena información. Para navegar a través de la Web es necesario un programa llamado navegador que envía peticiones de información a los servidores e interpreta, en forma de textos, gráficos y sonidos, las páginas que el servidor entrega como respuesta.

El funcionamiento de la Web se basa de este modo en la estructura hipertextual de sus contenidos. El hipertexto permite a los autores de una página enlazar sus documentos con otros documentos relacionados que se almacenan en otros servidores. De esta forma, los usuarios de la Web pueden navegar de unos servidores a otros a través de contenidos hipermediales que utilizan la tecnología del hipertexto.

El modelo cliente-servidor sobre el que se construye la Web, unido a la utilización de la tecnología del hipertexto, dan como resultado una Web con una estructura descentralizada y distribuida. Este tipo de red permitió el desarrollo de herramientas capaces de navegar a través de ella extrayendo, analizando y organizando su contenido. Estas herramientas son los denominados web crawlers.

Los web crawlers son casi tan antiguos como la web. En 1993, Matthew Gray desarrolló el World Wide Web Wanderer para obtener estadísticas sobre la evolución de la web. Las páginas recuperadas eran almacenadas e indexadas en el índice

Wandex, dando lugar así al primer buscador de la Web. Poco después surgirían otros ejemplos como el World Wide Web Worm (McBryan, 1994) o el Web Crawler (Pinkerton, 1994). En los años posteriores aparecerían los primeros buscadores comerciales (Lycos, Infoseek, AltaVista), que ya indexaban millones de páginas en sus índices.

Los web crawlers nacen en el ámbito de la Ingeniería Informática, vinculados con el área de la Recuperación de Información (*Information Retrieval*). La Recuperación de Información (RI) es un campo de estudio que tiene por objetivo encontrar contenidos (normalmente documentos), de naturaleza no estructurada (normalmente texto) que satisfacen una necesidad de información entre una gran colección de contenidos (normalmente almacenada en ordenadores).

Como señalan Abadal y Codina (2005), de las operaciones propias de la RI, la más característica es la *selección* de documentos, en función de criterios variables. Las características de su *contenido*, (los temas tratados), su *contexto* (p.e. la fecha de publicación,) o alguna combinación de ambas cosas (p.e: "documentos sobre desarrollo humano publicados por UNESCO entre 2003 y 2005").

Los sistemas de RI presentan algunos rasgos particulares en relación a otros tipos de procesamiento. Nuevamente en palabras de Abadal y Codina (2005).

1. *Uso de ordenadores (automatización)*. La RI se caracteriza por el uso de ordenadores y, por tanto, por el uso de bases de datos u otros sistemas automáticos o semiautomáticos de procesamiento de la información, tales como hipertextos. Aunque es lógicamente posible desarrollar sistemas de RI exclusivamente manuales, la teoría (y la práctica) de la RI nació de hecho con las primeras bases de datos y la mayoría de sus procedimientos o algoritmos sólo tienen sentido en un medio automatizado.

2. *Uso de información textual*. La RI gestiona información textual de tipo narrativo o discursivo, en lugar de, por ejemplo, datos numéricos o alfanuméricos muy estructurados, como hacen otros sistemas de información, por ejemplo, los sistemas administrativos (Salton; McGill, 1983: viii). Cuando la RI gestiona documentos u objetos no textuales, como imágenes, fotografías, video, etc., lo hace también a través de descripciones textuales (p.e., descripciones de las

imágenes) y/o de conjuntos de palabras que expresan el contenido y el contexto de las imágenes.

3. *Contexto de descubrimiento*. La RI se caracteriza por tener lugar en un contexto en el cual los usuarios del sistema de información tienen la necesidad de descubrir qué entidades cumplen una o más condiciones, , por ejemplo, qué documentos contienen información relevante para interpretar, desde el punto de vista x , el tema y . En otros sistemas de información, en cambio, los usuarios, partiendo de una entidad previamente conocida, quieren saber algo más de ella. La diferencia entre *descubrir cosas* y *ampliar datos* es esencial para entender la naturaleza de la RI.

Un buen sistema de RI debe ser capaz de responder a preguntas de descubrimiento de información en un entorno en el cual no es trivial el hecho de acceder a los documentos por su contenido.

Con el nacimiento de la Web, se desarrolla un área especializada vinculada con la Recuperación de Información dedicada al desarrollo de técnicas y herramientas para el estudio de la web: la Minería Web. Esta disciplina toma prestadas técnicas y metodologías del ámbito de la Minería de Datos, de ahí que adopte el nombre de Minería Web o *Web Mining*, y tiene por objetivo descubrir información útil a partir de la estructura de enlaces de la web, los contenidos de las páginas que la componen y los datos de uso de las mismas.

Tradicionalmente, se distingue la Recuperación de Información y la Minería de Datos por la diferente naturaleza de la información que estudian. Mientras que la Recuperación de Información estudia información de tipo textual, no estructurada o semiestructurada, la Minería de Datos estudia información de naturaleza no textual y estructurada.

El término “información estructurada” se refiere al tipo de información que tiene una clara estructura semántica lo que permite que sea fácilmente interpretada por un ordenador. El ejemplo típico de este tipo de información es el que contiene una base de datos relacional. Este es el tipo de información con el que generalmente trabaja la Minería de Datos. Otra característica de la información estructurada es que consiste en tipos de datos más o menos bien establecidos, como direcciones, datos de ventas, datos demográficos, etc.

A diferencia de este tipo de contenidos perfectamente descritos, la información no estructurada o semiestructurada es aquella que no tiene una estructura semántica explícita y por ello es más difícil de interpretar por un ordenador. Otra característica de la información no estructurada es que suele consistir en texto o discurso, en lugar de, por ejemplo, datos en hojas de cálculo. Este tipo de información necesita de una serie de procesos previos que permitan su correcta recuperación, análisis y organización. Los crawlers son herramientas que desarrollan esos procesos.

Para tratar de facilitar el acceso y recuperación de la información en Internet es que se creó el proyecto de web semántica. Su objetivo fundacional fue desarrollar una serie de tecnologías que permitieran a los ordenadores no solo entender el contenido de las páginas web y facilitar el proceso de gestión de la información en las búsquedas en la web, sino llevar a cabo razonamientos sobre las mismas.

Codina (2010: 5) señala los siguientes componentes conceptuales como principales características del proyecto de web semántica:

- **La web entendida como una gran base de datos:** la idea es conseguir que los documentos publicados en la web estén marcados de manera que sean similares a los registros de una base de datos
- **Metadatos:** las páginas web estarían caracterizadas con el uso intensivo de sistemas de metadatos como parte de su marcaje
- **Ontologías y lógica formal:** se desarrollarían ontologías para que los ordenadores interpreten la semántica de las páginas web, y sistemas de razonamiento basados en lógica formal que podrán hacer inferencias
- **Agentes de usuario:** Los sistemas informáticos serían capaces de representar los intereses de sus usuarios y de interactuar con otros sistemas sin la intervención de los usuarios.

Para llevar a cabo el proyecto de web semántica es imprescindible la aplicación de lenguajes controlados en la descripción de los contenidos de la web. Los lenguajes controlados son mecanismos utilizados para la presentación y organización del

conocimiento que tienen por objetivo controlar y normalizar la asignación de palabras clave a un documento (Vallez et al. 2010) y existen varios tipos

- **Taxonomías:** Son una forma de clasificación jerárquica del conocimiento que permite relacionar unidades de información organizándolas en clases y subclases.
- **Tesauros:** Son listas de términos usadas para organizar el conocimiento de un ámbito y controlar la descripción temática de un documento. Se basa en términos, descriptores y relaciones semánticas y funcionales que se establecen entre los términos. Estas relaciones pueden ser de equivalencia, asociación o jerarquía
- **Ontologías:** son la especificación del conocimiento de un determinado ámbito a través de la organización de conceptos, normalmente estructurados jerárquicamente, para representar entidades, ideas o eventos además de su propiedades y relaciones en un sistema de categorías.

Existen otras propuestas para la representación del conocimiento entre las que destacan las redes semánticas de Quillian (1968), los gráficos conceptuales de Sowa (1991), el análisis conceptual formal (Ganter y Wille, 1999), el marco de los espacios conceptuales (Gärdenfors y Williams, 2001), los mapas conceptuales (Novak y Cañas, 2006) o los mapas temáticos (Park y Hunting, 2003). En la actualidad, y en relación al proyecto de web semántica, parece poco aventurado decir que las ontologías en lenguaje SKOS y OWL integrado al lenguaje RDF, son las que mayor acogida están teniendo por parte de los desarrolladores, sin embargo, todavía siguen siendo minoritarias las páginas web que incorporan este tipo de tecnología.

Debido al interés por estudiar la información que contiene la Web, los crawlers fueron rápidamente adoptados por las Ciencias Sociales en general y por las Ciencias de la Información y la Comunicación Social en particular. En el ámbito de las Ciencias de la Información, la aplicación de crawlers y otras técnicas y herramientas relacionadas con la Recuperación de Información da lugar a una nueva área de conocimiento: la webometría.

A partir de concepciones y teorías propias de la bibliometría y la cienciometría, la webometría lleva a cabo estudios informétricos sobre la información contenida en la Web. Según Bjornebern y Ingwersen (2004: 1217) la webometría es “el estudio de los aspectos cuantitativos que intervienen en la construcción y uso de recursos de información, estructuras y tecnologías que componen la web desde perspectivas bibliométricas e informétricas”.

En palabras de Thelwall (2012), y ya con una mirada más alejada del ámbito estricto de las Ciencias de la Información, la webometría “es el estudio de información contenida en la web a través de métodos principalmente cuantitativos para lograr objetivos de investigación en el ámbito de las ciencias sociales utilizando técnicas que no son específicas de un campo de estudio”.

Como señalábamos en el Capítulo 1: Redes Temáticas en el ámbito de la Comunicación Social, el rastreo y análisis realizado con web crawlers se ha utilizado en numerosas líneas de investigación del ámbito de la Comunicación Social (Ackland, 2005; Park y Thelwall, 2008; Rogers y Ben-David, 2008; Shumate y Lipp, 2008; Shumate, 2012; Hsu y Park, 2012). El análisis de redes temáticas también se servirá de estas herramientas y puesto que comparte un proceso de trabajo común basado en la recuperación, análisis y organización de la información que contiene la web, también su marco teórico se construye principalmente a partir del ámbito de la Comunicación Social pero con elementos importantes del ámbito de las Ciencias de la Información y de la Ingeniería Informática.

Como vemos, el rastreo de la Web utilizando crawlers es un área con muchas aristas y no se puede dividir fácilmente en una clasificación de temas y subtemas. A modo de guía trataremos de responder tres preguntas básicas: ¿Qué son los web crawlers? ¿Para qué sirven? ¿Cómo funcionan?

Los web crawlers son programas que buscan, recuperan, analizan y en ocasiones almacenan los documentos de la Web. De la misma manera que un usuario puede navegar de una página web a otra a través de los hiperenlaces, los web crawlers utilizan los hiperenlaces para moverse a través de la Web visitando páginas y analizando la información que contienen para después almacenarla o indexarla en un índice o repositorio.

El proceso de rastreo de un crawler se inicia a partir de un conjunto de semillas o direcciones web (URLs) que pueden ser dadas por el usuario o por otro programa. Este conjunto inicial de URLs o semillas funciona como frontera inicial del crawler. La frontera de un crawler es la lista de URLs que debe visitar y esta se va modificando a medida que el crawler lleva a cabo el rastreo de la web. En cada ciclo de trabajo, el crawler elige una URL de la frontera, recupera la página y extrae los enlaces que contiene para añadirlos a la frontera. En función de las necesidades copiará la página en un servicio propio de almacenamiento o simplemente la indexará y continuará con el rastreo de las URLs hasta lograr el número de páginas suficientes que le permitan alcanzar el objetivo marcado.

En el ciclo de trabajo de un crawler existen diferentes fases que se pueden combinar de numerosas formas: 1) fetching o recuperación de la página web, 2) parsing o recuperación del contenido de la página, 3) preprocesamiento de los enlaces y textos recuperados, 4) indexación y/o almacenamiento de las páginas recuperadas. Vamos a considerar estas actividades con más detalle.

- **Fetching** o recuperación de la página web: para recuperar una página web el crawler actúa como un programa cliente haciendo una petición HTTP al servidor y leyendo la respuesta en busca de redirecciones e información sobre el estado de la página.
- **Parsing** o recuperación del contenido de la página: una vez descargada la página el crawler analiza su contenido y extrae la información necesaria para alimentar el siguiente proceso de trabajo. Esta información consiste típicamente en los enlaces que contiene la página pero puede incorporar otro tipo de informaciones etiquetadas en el código HTML de la página o en el DOM (Document Object Model)⁹.
- **Preprocesamiento** de enlaces y textos: los enlaces que se recuperan de la página deben ser transformados en un formato estándar o canónico que asegure el correcto procesamiento por parte del crawler una vez han sido

⁹ El DOM (Documento Object Model) de una página web escrita en HTML representa la estructura de la misma como un árbol de etiquetas que es incorporan al código de la página.

incorporados a la lista de URLs o frontera. Cuando se recupera el contenido de una página suele ser necesario eliminar las palabras vacías de significado o stopwords, estas son preposiciones, conjunciones, artículos, etc. Otro tipo de procesamiento de los textos que se suele llevar a cabo es el de stemming o lematización que consiste en reducir las palabras a su raíz significativa eliminando las desinencias morfológicas.

- **Indexación y almacenamiento:** las páginas recuperadas se indexan y se almacenan parcial o totalmente en un repositorio en función de los objetivos del crawler. Las páginas pueden ser almacenadas como documentos independientes o combinarse para ser almacenadas en un único archivo.

En cada una de estas fases del proceso de trabajo de un crawler intervienen algoritmos que pueden variarse en función de las necesidades pero la base de su funcionamiento será siempre similar a lo representado en la Figura 13 que vemos en la siguiente página.

En función de las características de su funcionamiento los crawlers se suelen clasificar en dos grandes grupos: crawlers no preferenciales y crawlers preferenciales. La diferencia entre unos y otros está en que los crawlers no preferenciales recuperan páginas de la web sin tener en cuenta ningún criterio de selección. Simplemente recuperan todas las páginas web que alcanzan a través de los enlaces. Los crawlers preferenciales en cambio, son selectivos y se basan en criterios como la actualización, el tamaño o el número de enlaces que recibe un sitio para seleccionar las páginas web que recuperan.

En la Figura 13 vemos el esquema de funcionamiento de un crawler de tipo no preferencial, concretamente de uno de tipo *Breadth-First-Crawler*. Este tipo de crawler se caracteriza por que en su lista de URLs o frontera no existe otra prioridad que el orden en que se recuperan las URLs a medida que el crawler rastrea páginas web. Este tipo de lista se conoce como FIFO *queue* (*first-in-first-out*) y en ella la primera URL a analizar es la que está ubicada en el primer lugar de la lista mientras que las URLs recuperadas se van añadiendo en el último lugar de la fila sin atender a otros criterios de ordenación diferentes al orden en que se recuperan los enlaces de las páginas web.

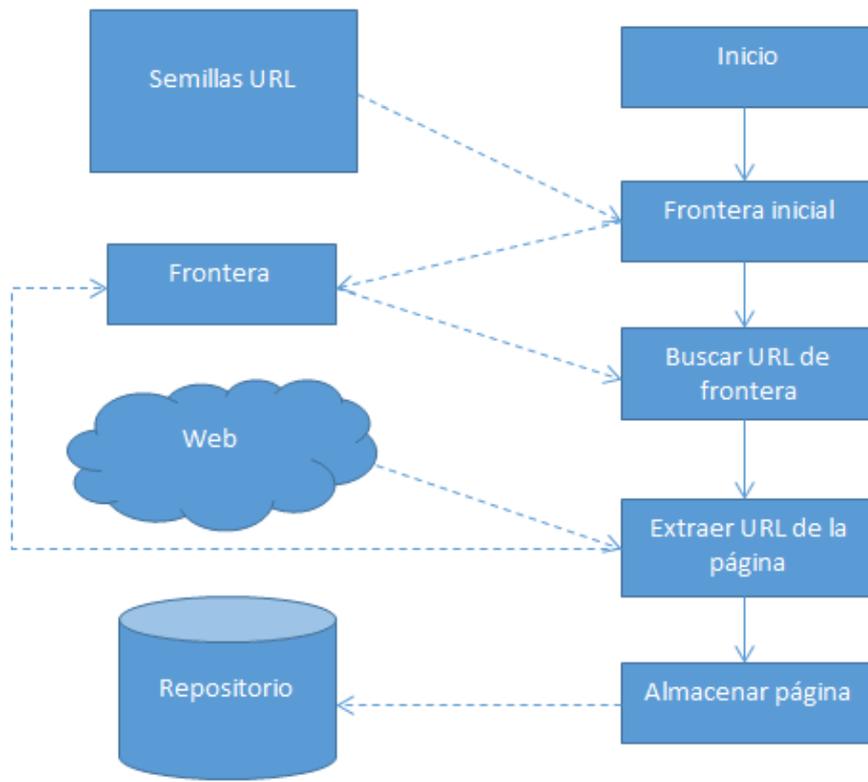


Figura 13. Esquema tipo de un crawler. Elaboración propia.

Que el crawler no utilice ningún criterio a la hora de visitar las páginas y ordenar los resultados no significa que el rastreo siga un orden aleatorio. La estructura de la web no es la de una red aleatoria sino la de una red libre de escala con una distribución de los enlaces que corresponde a una función de ley de potencia. La estructura de la web determinará el funcionamiento del crawler como vemos a continuación en la Figura 14.

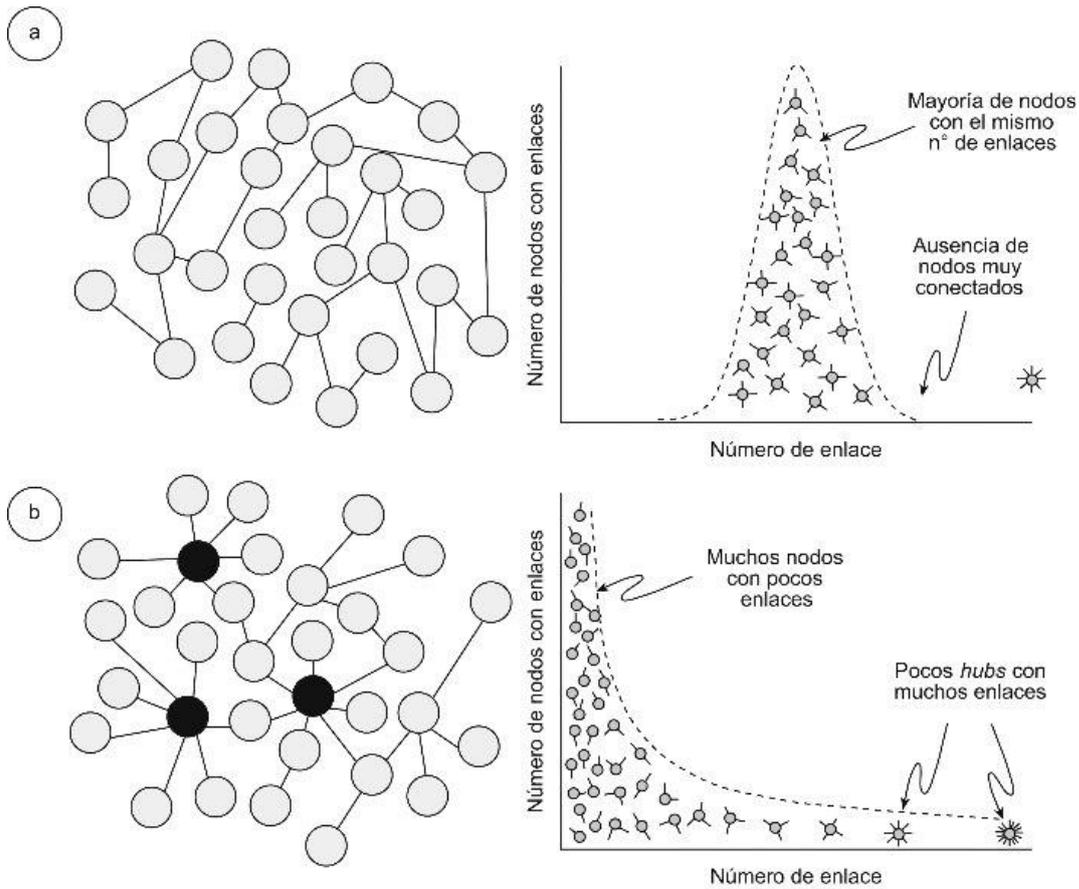


Figura 14. Representación gráfica de una red aleatoria y una red libre de escala junto a las funciones numéricas de nodos y enlaces (Perpiñá, 2009).

En la Figura 14a se representa una red aleatoria en la que la distribución de los enlaces corresponde con una distribución de Poisson, lo que significa que la mayoría de nodos tiene el mismo número de enlaces y hay muy pocos nodos que tengan muchos o muy pocos enlaces.

En la Figura 14b se representa una red libre de escala en la que el número de enlaces de unos pocos nodos es muy alto mientras que la mayoría de los nodos tienen un número muy bajo de enlaces. Esto es lo que se conoce como *long tail* (larga cola), por la forma de la función que lo representa, y que es característico de todas las redes libre de escala, entre ellas la web.

En la web se produce lo que se conoce como el Principio de Pareto, esto es que el número de enlaces de la web se divide en proporciones de 80-20, es decir, el 80 % de

los sitios acumula el 20 % de los enlaces mientras que el 20 % de los sitios acumula el 80 % de los enlaces.

En la web se produce además, y debido a las dos características que acabamos de señalar, el efecto *rich get richer*, que como ya señalábamos en el Capítulo 1, supone que las páginas más enlazadas de la web tienden a atraer los enlaces de las nuevas páginas que se incorporan a la Web. Lo mismo sucede con los crawlers no preferenciales del tipo *breadth-first*. Por las razones que hemos señalado, los crawlers se verán atraídos hacia esas páginas más conectadas que en el ámbito de la Web se conocen como *hubs*. Por la distribución de ley de potencia de los enlaces de la Web, un crawler que no utilice ningún criterio de prioridad en el análisis de enlaces se verá indefectiblemente atraído hacia los *hubs* que acumulan ese 80 % de enlaces que señala el principio de Pareto.

Otra razón por la que el rastreo que llevan a cabo los crawlers no preferenciales no se puede considerar aleatorio es la elección de las semillas iniciales. La probabilidad de que las páginas que se utilizan como semillas estén relacionadas con otras topológicamente cercanas es mucho mayor que la probabilidad de que estén enlazadas con otras páginas escogidas aleatoriamente.

Entre los crawlers no preferenciales también se encuentran los crawlers universales. Estos son los que utilizan los grandes buscadores y a diferencia de los *breadth-first* crawlers necesitan rastrear la web en su totalidad y de manera periódica. La arquitectura de este tipo de crawlers permite el desarrollo de procesos de trabajo paralelos en un mismo crawler o incluso coordinar eficazmente el trabajo de varios crawlers en varias máquinas diferentes. Todo ello tratando siempre de mantener un equilibrio entre la cobertura, la actualización y la relevancia de las páginas que recuperan.

Este tipo de crawlers se enfrenta además al problema que supone la estructura en forma de pajarita que tiene la web (*bow tie model*). Algunos estudios sobre la estructura estática de la web desvelaron que la web en su conjunto está compuesta por un grupo central de sitios fuertemente conectado, un segundo grupo que se conecta con el grupo central pero que no puede ser alcanzado desde él, y un tercer grupo que puede ser alcanzado desde el grupo central pero que no puede alcanzar a

este. Además de estos tres grupos principales existen otros grupos débilmente conectados o completamente desconectados (Broder et al., 2000). Lo vemos en la Figura 10

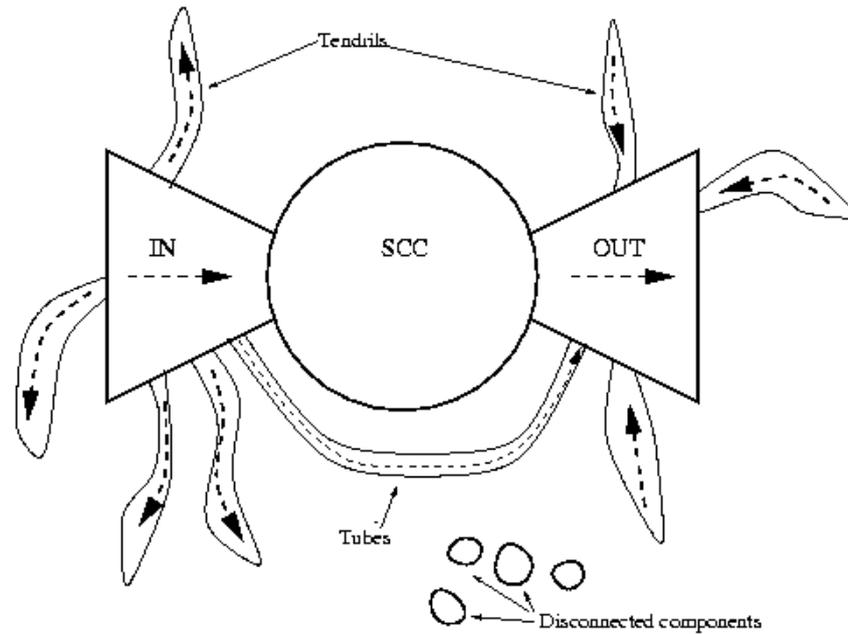


Figura 15. Representación gráfica de la estructura de la web según el modelo de pajarita (Broder et al, 2000).

Este tipo de estructura provoca que existan páginas web que no pueden ser alcanzadas a partir de los enlaces desde un punto inicial, y por tanto, que los crawlers universales dejen fuera de su rastreo amplias zonas de la web. Estas zonas forman parte de lo que actualmente se conoce como Web profunda o Deep Web y se requiere de crawlers con configuraciones especiales para alcanzarlas.

Además de las dificultades que supone la estructura de la web para el rastreo de los crawlers universales, estos también ven dificultada su labor por el carácter temporal de sus contenidos. Uno de los objetivos de este tipo de crawlers es mantener un índice de contenidos actualizados lo que choca con el alto nivel de modificación de los contenidos y de aparición y desaparición de sitios en la web.

Algunos estudios (Dasgupta et al., 2007; Ntoulas et al., 2004) han analizado la creación y eliminación de sitios en la web y la modificación de sus contenidos y de los enlaces que albergan aportando los siguientes resultados.

- Se crean nuevas páginas a un ritmo del 8 % de crecimiento semanal.
- Se eliminan páginas a un ritmo del 80 % anual.
- Se crean enlaces nuevos a un ritmo del 25 % semanal.
- Se eliminan enlaces a un ritmo del 80 % anual.

A pesar de esto, las web tiene la característica de ser una red de mundo pequeño¹⁰ y se estima que es posible alcanzar el 90 % de las páginas que contiene la web a partir de un grupo bien elegido de cinco o incluso menos URLs.

A diferencia de lo señalado con los crawlers no preferenciales, los crawlers preferenciales utilizan una lista de URLs o frontera en la que se otorga un orden de prioridad a unas URLs sobre otras en base a criterios variables. En cada ciclo de trabajo de un crawler preferencial se escoge la URL con mayor prioridad en la lista frontera, se recupera la página web y las URLs de los enlaces que contiene son añadidos a la lista frontera previa adjudicación de un valor de prioridad basado como decíamos en diferentes criterios de ordenación como el tamaño de los sitios, la actualización o el número de enlaces que reciben.

Entre los crawlers preferenciales, los más desarrollados son los focused crawlers o topical crawlers. Algunos autores diferencian unos de otros (Liu, 2007) mientras que otros prefieren verlos como diferentes evoluciones del mismo tipo de crawler (Olston y Najork, 2010). Sea cual sea la clasificación que hagamos estos se caracterizan por dirigir el rastreo hacia páginas que se consideran más probablemente relacionadas con las categorías en que el usuario está interesado.

En la Tabla 8 que extraemos de Olston y Najork (2010) se analizan los diferentes criterios de selección que utilizan los crawlers y cómo se relacionan con diferentes objetivos del rastreo web.

¹⁰ Se conoce como red de mundo pequeño una red en la que dos objetos pueden conectarse con pocos pasos intermedios.

Para localizar los contenidos relevantes para el usuario los focused crawlers utilizan información de contexto que guía la navegación a través de los enlaces. Esta información de contexto puede ser de tres tipos: contexto del enlace, páginas enlazadas y gráfico web (Pants & Srinivasan, 2005).

Technique	Objectives		Factors considered		
	<i>Coverage</i>	<i>Freshness</i>	<i>Importance</i>	<i>Relevance</i>	<i>Dynamicity</i>
Breadth-first search [43, 95, 108]	✓				
Prioritize by indegree [43]	✓		✓		
Prioritize by PageRank [43, 45]	✓		✓		
Prioritize by site size [9]	✓		✓		
Prioritize by spawning rate [48]	✓				✓
Prioritize by search impact [104]	✓		✓	✓	
Scoped crawling (Section 4.2)	✓			✓	
Minimize obsolescence [41, 46]		✓	✓		✓
Minimize age [41]		✓	✓		✓
Minimize incorrect content [99]		✓	✓		✓
Minimize embarrassment [115]		✓	✓	✓	✓
Maximize search impact [103]		✓	✓	✓	✓
Update capture (Section 5.2)		✓		✓	✓
WebFountain [56]	✓	✓			✓
OPIC [1]	✓	✓	✓		

Tabla 8. Lista de técnicas, objetivos y criterios de selección utilizados en los crawlers (Olston y Najork, 2010)

- El **contexto del enlace** se refiere al contenido léxico que se encuentra relacionado con el enlace; este puede limitarse al texto del enlace o a todo el contenido de la página que contiene el enlace.
- Utilizar las **páginas** enlazadas como información de contexto supone tener en cuenta el contenido léxico del texto de los enlaces que recibe una página o de las páginas que envían sus enlaces hacia la página a analizar.
- El **gráfico web** como información de contexto supone analizar el gráfico que se forma a partir de los enlaces entrantes y salientes de la página a analizar.

Los métodos y algoritmos que se utilizan para asignar un valor de similitud o relevancia a las URLs son varios pero suelen utilizar clasificadores basados en

métodos supervisados de aprendizaje (supervised learning) como Naïve Bayes, frecuencias de palabras (TD-IDF) o espacios vectoriales¹¹.

Tanto en los métodos supervisados de aprendizaje como en los no supervisados existen numerosos algoritmos y técnicas pero la variedad es tan grande que no nos permite estudiarlos en profundidad en esta investigación. Nos limitaremos a evaluar su implementación junto con otras características en las diferentes herramientas de análisis de redes temáticas que presentaremos en el siguiente apartado.

Hemos tratado de presentar una visión de conjunto de las características y funcionamiento de los crawlers pero no hemos visto en ella consideraciones particulares relacionadas con el análisis de información en la web semántica. ¿Cómo se adaptan los crawlers a las características y potencialidades de la web semántica? Seleccionamos algunas ideas planteadas por Blázquez (2014) en este sentido que comentamos brevemente a continuación.

Según Blázquez (2014: 125), existen tres métodos para construir la web semántica. Las especificaciones oficiales de RDF especifican que el método natural para desarrollar la web semántica es la codificación de un archivo en formato RDF y vinculado a la página web. Un segundo enfoque, el RDF-A, hace posible integrar la capa semántica en una página web utilizando el lenguaje XHTML. Una tercera opción es el empleo de microformatos que actúan directamente en el código HTML de las páginas.

Naturalmente los tres métodos ofrecen una serie de ventajas e inconvenientes y el método más adecuado será aquel que permita establecer relaciones semánticas entre los contenidos de los sitios web de una forma simple para los desarrolladores de páginas web y que al mismo tiempo sean fácilmente interpretables por los crawlers. Como señala Blázquez (2014: 125): “si se pretende divulgar el uso de la web semántica, su método de codificación debería aspirar a ser tan sencillo como el del propio lenguaje HTML”.

¹¹ Para más información sugerimos consultar los Capítulos 3 y 4 de Liu, B. (2007). *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media.

Los web crawler emplean métodos de análisis que actúan sobre el código fuente de las páginas web. El reconocimiento semántico, implica la utilización de nuevos métodos de detección de estructuras RDF y recolección de atributos. Una de las soluciones empleadas hasta ahora es la integración DOM (modelo de objetos del documento), que permite organizar jerárquicamente las etiquetas de un documento tanto en formato HTML como XML.

Los sistema de procesamiento y almacenaje deberán igualmente modificarse para permitir que el crawler indexe y almacene la información recuperada sin perder la información semántica que contiene. Para conseguirlo es necesaria la construcción de un BSD o *big semantic document*, que permita organizar los triples semánticos de las páginas web analizadas y así hacer posible su posterior recuperación a través de motores de búsqueda con capacidad para hacer inferencias semánticas y ofrecer sugerencias adicionales de información.

3.2. Caracterización y evaluación de herramientas para el análisis de redes temáticas en la web

En el apartado 3.1 hemos estudiado el marco teórico de los web crawlers, sus tipos, sus características y sus funcionalidades. En este apartado 3.2 nos proponemos describir y caracterizar los principales crawlers destinados a la investigación científica desde la perspectiva de un usuario investigador.

Para ello, en la sección 3.2.1. Perspectivas y modelos de calidad en la evaluación de software, estudiaremos el marco teórico relacionado con la evaluación de calidad en el software. En la sección 3.2.2. Selección y caracterización de herramientas, describiremos las características y funcionalidades de las herramientas seleccionadas y finalmente en el apartado 3.3. Propuesta de modelo de calidad y evaluación de herramientas, propondremos una herramienta de evaluación con una selección de parámetros, indicadores y métricas que nos permitan obtener una evaluación cuantitativa de las herramientas seleccionadas.

3.2.1. Perspectivas y modelos de calidad en la evaluación de software.

La calidad en el software ha sido una preocupación y uno de los principales objetivos de la ingeniería informática desde su nacimiento. Sin embargo, en un campo en constante evolución como la ingeniería informática, lo que significa calidad de software se ha convertido en una tarea difícil de delimitar.

La calidad en el software puede estudiarse desde numerosas perspectivas y no existe un consenso claro sobre las dimensiones que puede o debe abarcar el concepto de calidad en relación a un software. Probablemente, el primer antecedente en este sentido es la clasificación de Garvin que, aunque dirigida originalmente a la evaluación de la calidad del producto en general, ha servido de referente en el ámbito de la evaluación de la calidad en la ingeniería de software. Las cinco perspectivas propuestas por Garvin eran:

- Trascendente
- Basada en el producto
- Basada en el usuario
- Perspectiva de la producción
- Basada en el valor

Podemos ver esa influencia que comentábamos de la clasificación de Garvin en el ámbito de la ingeniería del software al estudiar la clasificación de Evans y Lindsay:

- **Subjetiva:** que no determina o define los parámetros por los que se mide la calidad del software. Calidad del producto: según la cual determinamos la calidad en función de la medición de una propiedad o atributo del software.
- **Calidad de la experiencia del usuario:** que tiene en cuenta la satisfacción que produce el software a un usuario para desarrollar una tarea.
- **Calidad del valor económico:** que establece una relación de utilidad-precio del producto software.

- **Calidad del proceso:** que evalúa el proceso de producción del software en base a ciertas especificaciones previas o normas de calidad de desarrollo de software.

Autores como Vaderonse, Ayed y Habra, (2015) simplifican la variedad de perspectivas contempladas por Garvin o Lindsay y Evans para reducirlas a dos grandes grupos: objetiva y subjetiva.

- La perspectiva de una **evaluación objetiva** es la que predomina en los modelos de calidad más importantes que han existido desde los orígenes de esta disciplina hasta el que se puede considerar el estándar de calidad más extendido actualmente, el ISO/IEC 25010. Según esta perspectiva, la calidad se puede evaluar a través de un conjunto predefinido de características que a través de la medición de ciertos datos del producto nos permitirá establecer un nivel de calidad relativa del producto respecto al patrón o modelo de calidad utilizado.
- La **perspectiva subjetiva**, sin embargo, define la calidad del software como un objetivo “líquido” y difícilmente reducible a un grupo de métricas. Según esta perspectiva, la calidad debe ser entendida de forma dinámica en conjunto con las experiencias de los usuarios y no únicamente a través de la aplicación de estándares predefinidos. Por tanto, el modelo de calidad se construye y puede evolucionar a través del tiempo en una especie de constructivismo continuado en relación con el cliente que da lugar a una perspectiva de la calidad que podríamos denominar emergente, holística o integral.

¿Cuántas veces nos hemos quejado al recibir las actualizaciones de un sistema operativo determinado; clamando al cielo que nos están vendiendo un software inacabado que solo gracias a la retroalimentación de los usuarios alcanza los requisitos mínimos de calidad? En realidad, este tipo de planteamiento, en el que se recaba la información de los usuarios para desarrollar un proceso continuo de mejoramiento del producto, responde a la lógica de un modelo de calidad ampliado que supera los modelos clásicos y supuestamente objetivos basados en la medición de un conjunto de características y especificaciones en un momento concreto.

A nuestro parecer, y apoyándonos en una revisión bibliográfica sistemática, son tres las perspectivas de la evaluación de calidad en software que más desarrollo están teniendo en el ámbito teórico y que más se están aplicando en el sector productivo: (1) software como producto, (2) software como proceso, (3) software como modelo.

- **Software como producto:** está relacionado con la perspectiva objetiva que señalábamos anteriormente, según la cual, la calidad se puede evaluar a través de un conjunto predefinido de características y a través de la medición de ciertos datos del producto software.
- **Software como proceso:** defiende que es posible mejorar la calidad final del producto software a través de la aplicación de modelos de calidad que controlen las diferentes fases del proceso de desarrollo de un software, por ejemplo su diseño, planificación, seguimiento y evaluación final. En algunos casos podemos encontrar que a este tipo de evaluación del software como proceso se le denomine evaluación de software como proyecto, reservando la denominación de evaluación como proceso para la que evalúa la calidad del software en función de los procesos productivos asociados.
- **Software como modelo:** surge de la aparición y desarrollo de la Ingeniería dirigida por modelos que presta más atención a la arquitectura del software que al código del software como tradicionalmente se había hecho.

Sin embargo, y en contra de la clasificación simplificada que acabamos de presentar, se están produciendo una serie de transformaciones en el ámbito de la ingeniería informática que hacen necesaria la revisión y ampliación del concepto de calidad del software. Según Breu, Kuntzmann-Combelles y Felderer (2015) son tres las principales transformaciones que se están desarrollando en la producción de sistemas informáticos y que provocan nuevas necesidades en el ámbito de la gestión de la calidad: la interconexión de servicios, la evolución de sistemas y la colaboración de los clientes.

Los sistemas informáticos consisten cada vez más de servicios fragmentados que se interconectan de manera descentralizada. Los flujos de trabajo comprenden habitualmente servicios móviles (*mobile computing*), servicios en la nube (*cloud*

computing) y servicios de sensores (M2M o *Machine to Machine computing*). En este contexto, la gestión de la calidad no puede limitarse a aspectos técnicos o de la plataforma en la que trabajan, sino que también deben tener en cuenta cuestiones relacionadas con la confianza de los usuarios, su seguridad y los servicios de atención externos que se les prestan.

Teniendo en cuenta las nuevas necesidades señaladas surgen una serie de propuestas de innovación que explicamos brevemente a continuación:

- **Gestión del conocimiento:** la generación, distribución y visualización de datos relacionados con el desempeño del software sirve de apoyo a los usuarios para el mejor cumplimiento de sus tareas y toma de decisiones además de facilitar el mejoramiento continuo de la calidad del software en relación con los desarrolladores del mismo.
- **Automatización:** relacionada con los modelos de arquitectura del software y la gestión continua de la calidad en sistemas informáticos evolutivos y adaptativos, la evaluación automática basada en modelos de software permite la evaluación inmediata del comportamiento del mismo, la detección temprana de riesgos y la generación de documentación en base a los test automáticos que se puedan desarrollar.
- **Análisis de datos:** comprende la aplicación de modelos de análisis de datos para la evaluación de la calidad en los diferentes procesos de un software, lo que nos permitirá determinar el estado actual del software, predecir estados futuros y mejorar así la toma de decisiones relacionadas con las tareas necesarias para mejorar la calidad del software.
- **Procesos de colaboración:** incluye todos los procesos automáticos, semiautomáticos y manuales de evaluación de un software relacionando a los usuarios y desarrolladores en un proceso conjunto de evaluación que incluya los datos obtenidos en el análisis del software en los diferentes niveles de la organización que lo desarrolla y utiliza.

Por las características y objetivos de esta investigación utilizaremos la perspectiva de la calidad del software como producto. Sin embargo, dentro de esta perspectiva de evaluación también existen diferentes tipos de modelos de calidad que podemos agrupar en tres categorías:

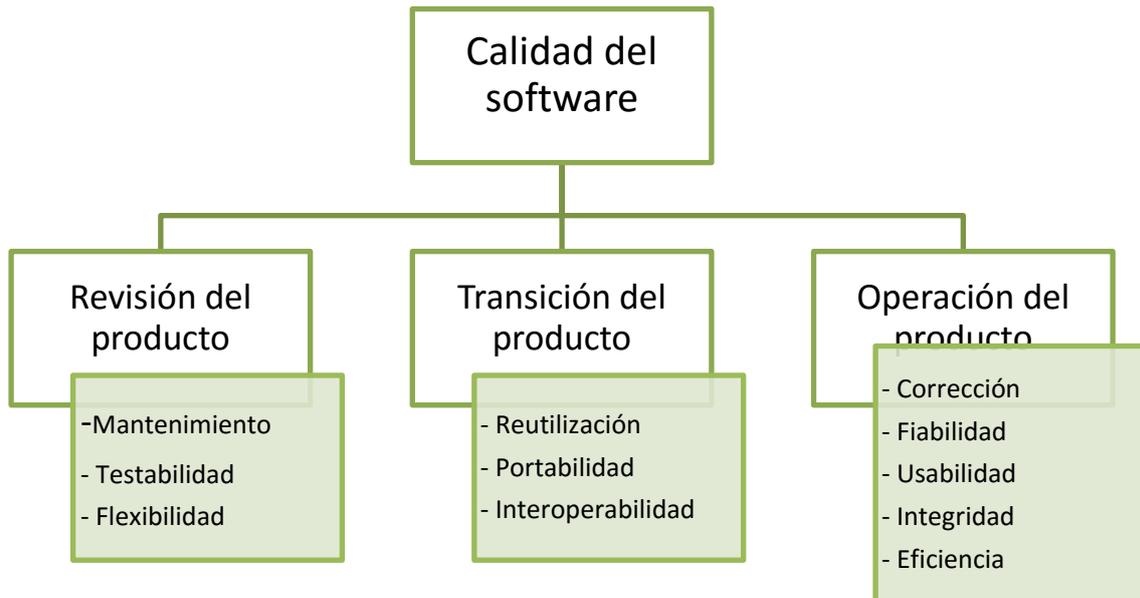
- **Jerárquicos:** descomponen el concepto de calidad en una estructura jerárquica de características y subcaracterísticas relacionadas con la calidad. Los modelos de McCall, Bohem o las normas ISO/IEC 9126 o 25000 son ejemplos de este tipo de modelo de calidad.
- **Metamodelos:** modelos del modelo de calidad que describen por ejemplo la validez de la estructura de estos modelos. El modelo de Dromey es probablemente el ejemplo más conocido de este tipo de metamodelos.
- **Modelos implícitos o estadísticos:** capturan propiedades del producto y estiman o predicen su calidad en base a métodos estadísticos.

Entre estos modelos analizaremos con más detalle los modelos jerárquicos y concretamente el modelo 25000 SQuaRE que es el estándar internacional más utilizado en la actualidad. Pero antes, consideramos que puede ser explicativo para el posterior proceso de evaluación presentar brevemente las características y evolución de los modelos de calidad a lo largo de la historia.

La primera propuesta de un modelo de calidad del software es la que desarrolla McCall en 1977 a petición de la US Air-Force Electronic System Decision (ESD), el Rome Air Development Center (RADC) y General Electric para mejorar la calidad de sus productos de software. El modelo de McCall evaluaba el producto software desde tres perspectivas: (a) revisión del producto, (b) transición del producto y (c) operación del producto. Dando lugar a lo que se ha llamado “El Triángulo de la Calidad de McCall”. Cada una de estas tres perspectivas está asociada con un conjunto de parámetros, factores según la terminología de McCall. Lo vemos a continuación en la Figura 16.

Cada uno de estos parámetros cuenta con una serie de indicadores que asociados a un conjunto de métricas nos permitirán obtener una evaluación cuantitativa de la

calidad del producto software. La principal contribución de este modelo fue precisamente la de relacionar características de calidad con métricas (Ortega, 2003; AL-Badareen et al., 2011) y, sin embargo, podemos ver que las perspectivas propuestas por McCall no corresponden con las que se han señalado como principales perspectivas utilizadas actualmente. A continuación presentamos una lista con los



parámetros y sus indicadores asociados. Como veremos, un mismo indicador puede ser utilizado para definir varios parámetros de evaluación.

Figura 16. Triángulo de la calidad de McCall (1977). Elaboración propia.

El siguiente modelo es el de Boehm (1978), que se puede considerar una evolución del modelo de McCall. El modelo de Boehm añade algunas características nuevas como la evaluación de la utilidad del producto y presta mayor atención al mantenimiento del software. Igual que sucedía en el modelo de McCall, Boehm propone un modelo jerárquico que incluye características de alto nivel, lo que McCall denominaba factores, características de nivel medio, que podríamos equiparar con los criterios propuestos por McCall, y finalmente un conjunto de métricas. Lo vemos en la Figura 17.

El modelo FURPS fue propuesto por Grady y Caswell (1987) de Hewlett-Packard Co. y tiene en cuenta cinco características que son las que le dan nombre por sus siglas en

inglés: functionality, usability, reliability, performance, supportability (funcionalidad, usabilidad, fiabilidad, rendimiento y compatibilidad).

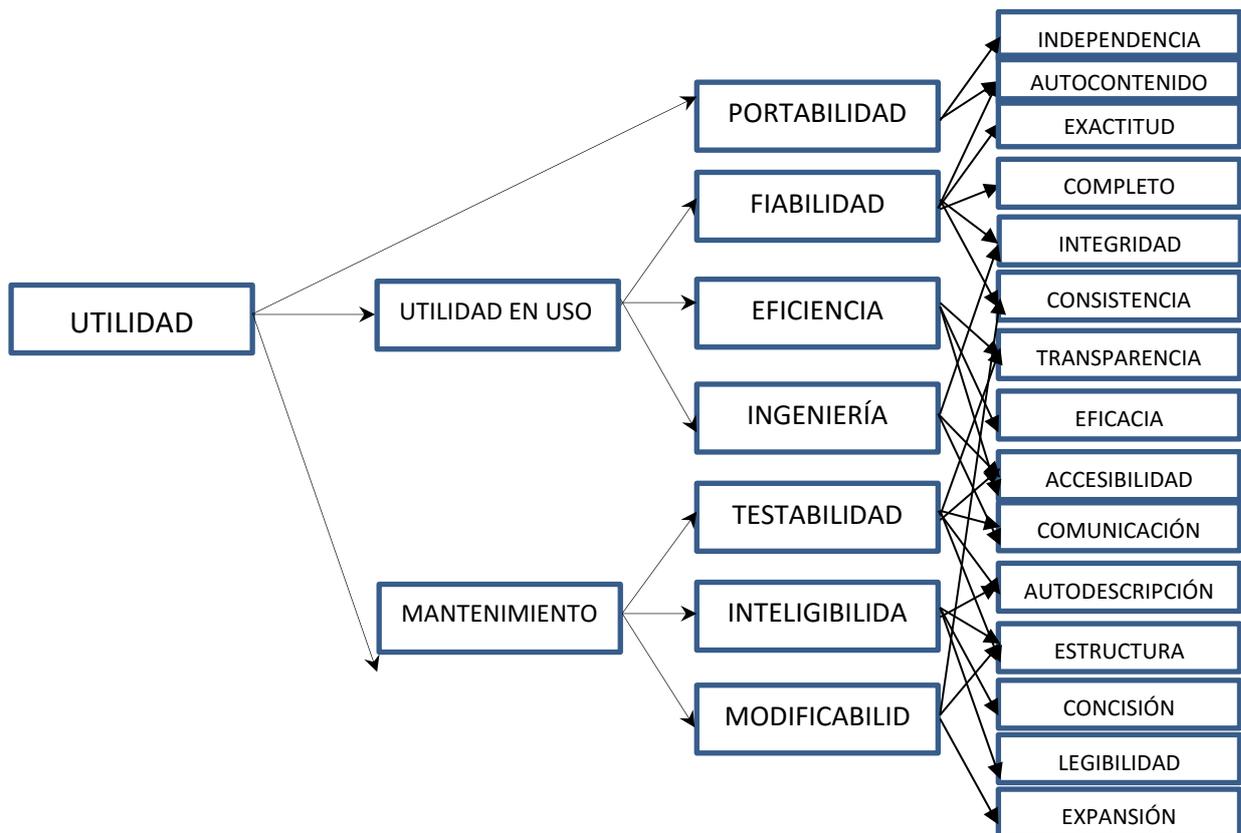


Figura 17. Modelo de calidad de Bohem, (1978.) Elaboración propia.

Estos cinco parámetros se clasifican en dos categorías: requisitos funcionales (F) y requisitos no funcionales (URPS). Este modelo considera únicamente las necesidades del usuario y no tiene en cuenta la evaluación de los desarrolladores. De esta forma, quedaban fuera del modelo parámetros habitualmente utilizados como la portabilidad o la facilidad de mantenimiento. Extraemos un detalle de cada uno de estos parámetros que traducimos de Eeles (2005). Los indicadores que se relacionan con el parámetro de la funcionalidad (F) del software son los siguientes:

Función	Descripción
Auditoría	Proporciona guías de auditoría de los sistemas de ejecución.
Licencia	Proporciona servicios de seguimiento, adquisición, instalación y monitorización de licencias de uso.
Localización	Proporciona facilidad para el soporte de múltiples lenguajes naturales.
Correo	Proporciona servicios que permiten aplicaciones de envío y recepción de mensajes.
Ayuda online	Proporciona servicios de ayuda online.
Impresión	Proporciona facilidades de impresión.
Reportes	Proporciona facilidades para la generación de informes.
Seguridad	Proporciona servicios para proteger el acceso a ciertos recursos de información.
Sistema de gestión	Proporciona servicios que facilitan la gestión de aplicación en entornos distribuidos.
Flujo de trabajo	Proporciona soporte para el traslado de documentos y procesos de trabajo incluyendo revisiones y ciclos de aprobación.

Tabla 9. Modelo de calidad FURPS, (Eeles, 2005). Elaboración propia.

Los parámetros restantes "URPS" describen requerimientos no funcionales que se relacionan con la arquitectura del software. A continuación hacemos una breve enumeración de los indicadores que se relacionan con cada uno de estos parámetros.

- **Usabilidad:** se relaciona con características como el aspecto estético, la consistencia o la interfaz del software.

- **Fiabilidad:** está relacionada con características como la precisión de los cálculos del sistema, la capacidad del software para reponerse a errores o la cantidad de recursos del sistema que deja disponible el software.

Rendimiento: tiene que ver con características como la cantidad de trabajo que desempeña el software, el tiempo de respuesta, el tiempo de recuperación o el tiempo de arranque y apagado del software.

- **Soporte:** está relacionado con características como la capacidad de testeado del software, la adaptabilidad, la compatibilidad, la facilidad de configuración, la posibilidad de ampliado o la localización.

Posteriormente, el modelo FURPS fue ampliado y mejorado incorporando nuevos parámetros que dieron lugar al modelo FURPS+ (Grady, 1992). Los parámetros incorporados fueron diseño, implementación, interfaz y requerimientos físicos.

- **Diseño:** especifica o limita las opciones para diseñar un sistema como puede ser la necesidad de utilizar una base de datos como núcleo del sistema.
- **Implementación:** especifica la codificación o construcción que debe utilizar el sistema. Por ejemplo la necesidad de utilizar estándares, lenguajes determinados o un límite de recursos.
- **Interfaz:** especifica los elementos del sistema a través de los cuales el usuario interactúa con el software, como pueden ser formatos u otros factores que intervienen en el proceso de interacción.
- **Requerimientos físicos:** hace referencia a los elementos y características que debe tener el hardware para albergar el sistema software.

Con estas modificaciones, el modelo cambia su perspectiva inicial que tenía en cuenta únicamente al usuario e incorpora la perspectiva del desarrollo del software.

A diferencia de los modelos jerárquicos, el modelo de Dromey (1995), y elegimos este como muestra de metamodelo, utiliza tres entidades primarias para construir su

modelo de calidad: un conjunto de componentes, un conjunto de propiedades de los componentes relacionadas con la calidad y un conjunto de atributos generales de calidad. Estas tres entidades de evaluación pueden relacionarse dando lugar a seis tipos de relaciones. Lo vemos en la siguiente figura.

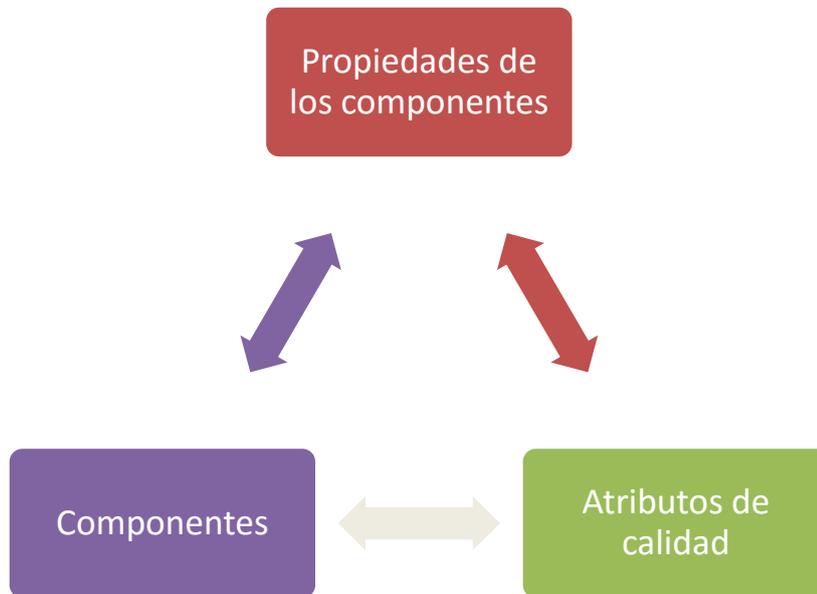


Figura 18. Modelo genérico de calidad de Dromey (1995). Elaboración propia.

Como señala el propio Dromey, para construir la calidad de un producto software o del proceso de desarrollo de ese software, solo son importantes cuatro de las seis posibles relaciones que propone el modelo. Estas son las relaciones superiores del gráfico, las que se representan con las flechas naranjas.

El modelo de Dromey tiene por objetivo ofrecer un sistema de evaluación flexible y adaptable a diferentes tipos de software y a diferentes perspectivas de la evaluación de calidad. Pensando especialmente en el trabajo de los desarrolladores de los programas. Para ello, la evaluación puede aplicarse siguiendo el modelo “de abajo hacia arriba”, es decir, al identificar qué propiedades de calidad tienen un impacto importante en qué atributos superiores de calidad, o “de arriba hacia abajo”, es decir, qué atributos superiores de calidad necesitan considerar unas u otras propiedades de calidad. Veamos esto de manera gráfica para tratar de aclarar las posibles relaciones del modelo:

Forma estructural¹² ———> Relaciones de las propiedades de calidad
 Propiedades de calidad ———> Relaciones de los atributos de calidad

Estas dos relaciones juntas nos permiten alcanzar la calidad de los parámetros superiores del modelo desde las propiedades particulares del producto. Siendo esta perspectiva especialmente útil para el proceso de desarrollo y los equipos de desarrolladores.

Atributo de calidad ———> Relaciones de las propiedades de calidad
 Propiedades de calidad ———> Relaciones de forma estructural

Estas dos relaciones nos permiten mejorar la calidad del producto software a partir de los atributos superiores del modelo desde “arriba hacia abajo” identificando qué propiedades necesitan satisfacerse utilizando cada forma estructural del sistema para a través de ellas alcanzar la calidad en cada uno de los atributos superiores de calidad. Esta perspectiva resulta especialmente útil para los diseñadores que tienen que establecer y operacionalizar los atributos de calidad en el diseño del software.

Una vez estudiados los que se pueden considerar principales modelos de calidad del software en la historia reciente de esta disciplina, pasamos a analizar con mayor profundidad el que probablemente es el modelo más utilizado y que más influencia ha tenido en el ámbito de la calidad del software: el estándar internacional de calidad ISO/IEC 9126. Esta norma de calidad fue introducida por primera vez en 1991 y desde entonces ha sido modificada en numerosas ocasiones hasta llegar a la última versión de 2001, y posteriormente ser sustituido en 2005 por el estándar actual ISO/IEC 25000. Este último, junto a otras normas, ya existentes algunas de ellas y otras en desarrollo actual ha dado lugar al denominado grupo SQuaRE, por las siglas en inglés de Software Quality Requirements Evaluation (Evaluación de requerimientos de calidad en el software).

¹² Dromey utiliza el concepto de “forma estructural” para hacer referencia a los tipos de declaración y los tipos de componentes del lenguaje de implementación del sistema.

Consideramos importante conocer las características principales del modelo ISO/IEC 9126 para comprender el actual modelo SQuaRE, lo que además nos permitirá ver la evolución y tendencias del ámbito de la evaluación de software. Por todo ello, a continuación procedemos a hacer un breve resumen de las características principales del modelo ISO/IEC 9126 para abordar posteriormente el modelo ISO/IEC 25000.

El modelo ISO/IEC 9126, al igual que el resto de modelos estudiados, se estructuraba jerárquicamente, en este caso, en cuatro niveles: (1) características de calidad, (2) sub-características, (3) atributos e (4) indicadores. En estos cuatro niveles se distinguen características internas (aspectos estáticos del producto de software) y características externas (aspectos dinámicos del producto de software) para completar el modelo con un grupo de características relacionadas con la perspectiva de la calidad en el uso o, dicho de otra forma, calidad desde el punto de vista del usuario final.



Figura 19. Modelo de calidad interior y exterior ISO/IEC 9126 (2001). Elaboración propia.

Las características internas y externas del software se descomponen en seis características que se definen a través de un grupo de subcaracterísticas. Lo vemos en la Figura 19.

En relación a la calidad en uso, o calidad desde la perspectiva del usuario final, el modelo se divide en cuatro características. Lo vemos a continuación en la Figura 20.



Figura 20. Modelo de calidad en uso ISO/IEC 9126 (2001). Elaboración propia.

Una de las peculiaridades del modelo ISO/IEC 9126 respecto a los modelos previamente estudiados es que diferencia entre las características internas y externas del software aunque sigue la misma lógica jerárquica que la mayoría de ellos. El modelo es fácilmente adaptable y modificable por lo que ha sido utilizado como base de otros modelos que evalúan aspectos más específicos del software (Vanderose, 2012).

El modelo de calidad ISO/IEC 9126 (calidad del producto software) se complementó posteriormente con la norma ISO/IEC 14598 (evaluación del producto de software). A pesar de pertenecer a un marco común, los ciclos de trabajo de ambas series generaban problemas en su aplicación lo que llevó a sustituirlas por el conjunto de normas ISO/IEC 25000 SQuaRE.

SQuaRE se compone de cinco series de normas: ISO/IEC 2500n-Quality Management Division, ISO/IEC 2501n-Quality Model Division, ISO/IEC 2502nQuality Measurement Division, ISO/IEC 2503n-Quality Requirements Division, and ISO/IEC 2504n-Quality Evaluation Division. A continuación presentamos una figura con una traducción libre de los nombres de las diferentes divisiones.



Figura 21. Divisiones del modelo ISO/IEC 25000 SQuaRE (2005). Elaboración propia.

La división de Gestión de la calidad se encarga de redactar las normas que definen todos los modelos, términos y definiciones comunes que se utilizan en el resto de normas de la familia 25000. Algunas de ellas se encuentran desactualizadas y siguen, por ejemplo, haciendo referencia a la división entre calidad interna y externa que existía en la ISO/IEC 9126.

- ISO/IEC 25000 - **Guide to SQuaRE**: contiene el modelo de la arquitectura de SQuaRE, la terminología de la familia, un resumen de las partes, los usuarios previstos y las partes asociadas, así como los modelos de referencia.
- ISO/IEC 25001 - **Planning and Management**: establece los requisitos y orientaciones para gestionar la evaluación y especificación de los requisitos del producto software.

El corazón de todo el conjunto de normas es el modelo de calidad que se define en la norma ISO/IEC 25010. Como ya hemos indicado previamente mantiene la estructura jerárquica propuesta en la ISO/IEC 9126 manteniendo incluso el doble modelo de calidad y calidad en uso del producto. Ambos modelos se modifican ligeramente incorporando algunas nuevas categorías en los diferentes niveles.

Las normas de este apartado presentan modelos de calidad detallados incluyendo características para calidad del producto software y la calidad en uso. Mientras que el modelo de calidad del producto software sirve para describir las características del producto directamente, la calidad en uso busca evaluar las características del software en relación con diferentes grupos de interés del producto, principalmente usuarios del mismo.

Actualmente esta división se encuentra formada por:

- ISO/IEC 25010 - **System and software quality models**: describe el modelo de calidad para el producto software y para la calidad en uso. Esta norma presenta las características y subcaracterísticas de calidad frente a las cuales evaluar el producto software..
- ISO/IEC 25012 - **Data Quality model**: define un modelo general para la calidad de los datos que se encuentran almacenados de manera estructurada y forman parte del sistema de información.

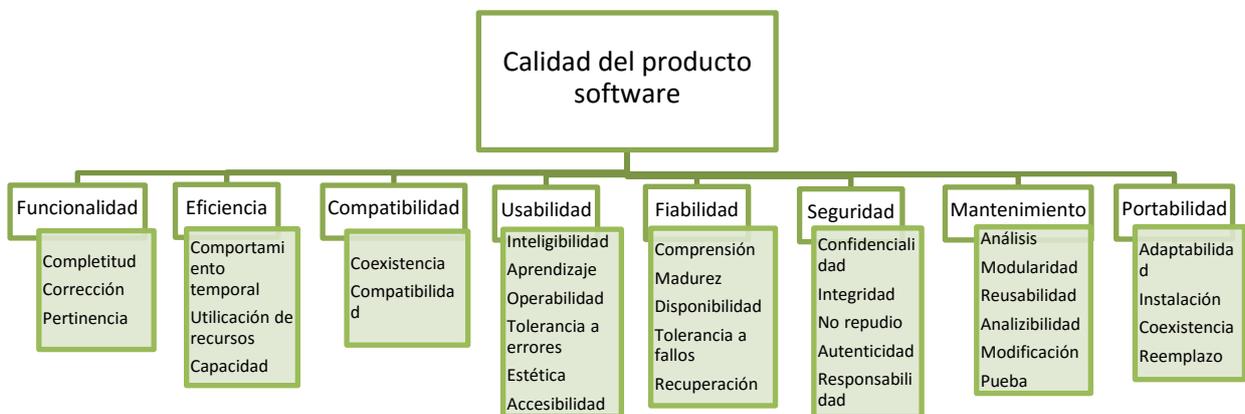


Figura 22. Modelo de calidad del producto software ISO/IEC 25010 SQuaRE (2011). Elaboración propia.

La División de Medición de Calidad se ocupa de crear las normas que sirven de modelo de referencia de la medición de la calidad del producto, definiciones de medidas de calidad (interna, externa y en uso) y guías prácticas para su aplicación. Actualmente esta división se encuentra formada por:



Figura 23. Modelo de calidad en uso ISO/IEC 25010 SQuaRE (2011). Elaboración propia.

- ISO/IEC 25020 - **Measurement reference model and guide**: presenta una explicación introductoria y un modelo de referencia común a los elementos de medición de la calidad. También proporciona una guía para que los usuarios seleccionen o desarrollen y apliquen medidas propuestas por normas ISO.
- ISO/IEC 25021 - **Quality measure elements**: define y especifica un conjunto recomendado de métricas base y derivadas que puedan ser usadas a lo largo de todo el ciclo de vida del desarrollo software.
- ISO/IEC 25022 - **Measurement of quality in use**: define específicamente las métricas para realizar la medición de la calidad en uso del producto.
- ISO/IEC 25023 - **Measurement of system and software product quality**: define específicamente las métricas para realizar la medición de la calidad de productos y sistemas software.
- ISO/IEC 25024 - **Measurement of data quality**: define específicamente las métricas para realizar la medición de la calidad de datos.

La División de Requisitos de Calidad genera las normas que ayudan a especificar requisitos de calidad que pueden ser utilizados en el proceso de elección de los

requisitos de calidad del producto software a desarrollar o como entrada del proceso de evaluación. Para ello, este apartado se compone de:

- ISO/IEC 25030 - **Quality requirements**: provee de un conjunto de recomendaciones para realizar la especificación de los requisitos de calidad del producto software.

La División de Evaluación de Calidad desarrolla normas que proporcionan requisitos, recomendaciones y guías para llevar a cabo el proceso de evaluación del producto software. Esta división se encuentra formada por:

- ISO/IEC 25040 - **Evaluation reference model and guide**: propone un modelo de referencia general para la evaluación, que considera las entradas al proceso de evaluación, las restricciones y los recursos necesarios para obtener las correspondientes salidas.
- ISO/IEC 25041 - **Evaluation guide for developers, acquirers and independent evaluators**: describe los requisitos y recomendaciones para la implementación práctica de la evaluación del producto software desde el punto de vista de los desarrolladores, de los adquirentes y de los evaluadores independientes.
- ISO/IEC 25042 - **Evaluation modules**: define lo que la Norma considera un módulo de evaluación y la documentación, estructura y contenido que se debe utilizar a la hora de definir uno de estos módulos.
- ISO/IEC 25045 - **Evaluation module for recoverability**: define un módulo para la evaluación de la subcaracterística Recuperabilidad (*Recoverability*).

Hay que entender que todos los modelos que hemos presentado son subjetivos y están sujetos a modificación en función de las necesidades de evaluación o las propias características del software a evaluar. En nuestro caso, el software para el análisis de redes temáticas en la web son mayoritariamente crawlers, con unas características muy específicas de este tipo de software. Por otro lado, los requerimientos que hacemos a este tipo de software también son específicos de

nuestra investigación y de nuestro perfil como investigadores del ámbito de la Comunicación Social y las Ciencias de la Información. En el apartado 3.3 propondremos un sistema de evaluación adaptado a esas características específicas de los crawlers que señalábamos y a los requerimientos propios de nuestra investigación.

3.2.2. Selección y caracterización de herramientas

Después de una revisión teórica exhaustiva de investigaciones que llevan a cabo análisis de redes de hiperenlaces en diferentes ámbitos de conocimiento (Barabasi, 2002; Brin y Page, 1998; Almind y Ingwersen, 1997) y análisis de redes temáticas en el ámbito de la Comunicación Social (Ackland, 2005; Park y Thelwall, 2008; Rogers y Ben-David, 2008; Shumate y Lipp, 2008; Shumate, 2012; Hsu y Park, 2012)., se han escogido cuatro crawlers como las herramientas que se pueden considerar más adecuadas para desarrollar este tipo de análisis: IssueCrawler, SocSciBot, Webometric Analyst y VOSON.

En este apartado nos proponemos describir las diferentes características de las herramientas seleccionadas para lo que aportaremos información sobre las instituciones que las desarrollan, las principales funcionalidades de cada una de ellas y una serie de imágenes que permitan visualizar el proceso de análisis que permiten llevar a cabo en relación con el análisis de redes temáticas en la web.

a) IssueCrawler

IssueCrawler¹³ fue desarrollado en 2001 por la Govcom.org Foundation de Amsterdam que dirige el investigador Richard Rogers. La Govcom.org se dedica al desarrollo y mantenimiento de herramientas de análisis web relacionadas con el ámbito de la política, siendo IssueCrawler la herramienta que más relevancia ha obtenido a nivel científico. Actualmente la fundación Govcom.org se ha incorporado a la iniciativa de

¹³ Disponible en <https://www.issuecrawler.net/>

investigación de la web Digital Methods Initiative¹⁴, en la que también participa la Universidad de Amsterdam y numerosas universidades y centros de investigación europeos.

IssueCrawler es un software de rastreo y visualización de redes de hiperenlaces accesible únicamente a través de la web, es decir, no permite su descarga e instalación en otro ordenador y para poder utilizarlo necesitamos una cuenta de usuario que debe ser previamente aceptada por el administrador del programa.

IssueCrawler consiste en un cosechador de enlaces (*harvester*)¹⁵, un crawler, un motor de análisis de co-enlaces y tres módulos de visualización. El programa está instalado en un servidor y lleva a cabo la recuperación y extracción de enlaces de sitios específicos (semillas) que debemos introducir como input en el programa. Sobre esos sitios iniciales o semillas puede realizar varios tipos de análisis de enlaces para identificar redes densamente interconectadas y visualizarlas en forma de mapas de clusters¹⁶. Originalmente, IssueCrawler ofrecía además la posibilidad de visualizar los datos a través de mapas circulares y mapas geográficos pero actualmente estas dos opciones no ofrecen resultados.

Los análisis que puede llevar a cabo IssueCrawler son de tres tipos aunque se pueden configurar diferentes aspectos en cada uno de ellos.

El análisis de co-enlaces (*co-link analysis*) rastrea las URLs introducidas como semillas e identifica los enlaces salientes de estas para recupera únicamente las páginas que reciben al menos dos enlaces por parte de las semillas. Posteriormente detecta los enlaces entrantes que reciben las semillas por parte de las páginas rastreadas en la fase inicial.

¹⁴ DMI (Digital Methods Initiative) es un grupo de investigación en Nuevos Medios que desarrolla además un programa de doctorado “New Media PhD”
<https://wiki.digitalmethods.net/Dmi/DmiAbout>

¹⁵ Un harvester es un programa que extrae URLs de un texto, un código fuente o de los resultados de un buscador produciendo una lista limpia de URLs.

¹⁶ Se denominan clusters los conjuntos de nodos altamente conectados de una red.

Instructions of use | Scenarios of use | Français | 한국어 | FAQ | Allied tools » Log Out

issuecrawler

the Lobby
Issue Crawler
Network Manager
Archive

Monday, August 24, 2015

Harvester

Type or paste text and URLs into the Harvester

The text will be stripped to create starting points for the Issue Crawler

Next step »
Harvest

@972.42

Current and Queued Crawls		RSS
25 Aug 2015	test 3	

Figura 24. Módulo harvester. Extraído de IssueCrawler.

El análisis en efecto bola de nieve (*snowball analysis*) identifica los enlaces salientes de las semillas para recuperar las páginas que reciben al menos un enlace por parte de estas.

El análisis de “Interactores” (*interactor analysis*) rastrea las URLs introducidas como semillas y analiza los enlaces recíprocos que existen entre las semillas introducidas.

Crawling Methods



Co-link
 Crawl seeds and retain URLs with at least two links from seeds. Co-linkees become seeds for subsequent iteration.



Snowball
 Crawl seeds and retain URLs with at least one link from seeds.



Inter-actor
 Crawl seeds and return inter-linkings among seeds.

Figura 25. Módulo de selección de rastreo. Extraído de IssueCrawler.

IssueCrawler permite además configurar cada uno de estos análisis en cuatro formas diferentes: Privilegio de las semillas (*Privilege Starting Points*), análisis por página o por sitio, número de iteraciones y profundidad del rastreo.

Privilegio de las semillas: nos permite decidir si queremos que se mantengan las semillas iniciales en los resultados del rastreo después de la primera iteración.

Análisis por página o por sitio: el análisis por página web analiza las páginas en profundidad y nos permite identificar redes de páginas. El análisis de sitios por el contrario analiza únicamente las páginas de inicio de un sitio web e identifica redes de enlaces entre sitios o páginas de inicio.

Número de iteraciones: nos permite definir el número de iteraciones que lleva a cabo el crawler con un límite máximo de tres iteraciones. En un rastreo de tipo bola de nieve nos permitirá analizar los enlaces salientes de las semillas (1ª iteración), los enlaces de las páginas que reciben un enlace de las semillas (2ª iteración), y los enlaces de las páginas que reciben un enlace de las páginas enlazadas en primer lugar por las semillas (3ª iteración).

Profundidad del rastreo: nos permite definir la profundidad del rastreo en un sitio web con una profundidad máxima de tres niveles. En el primer nivel el rastreo se limita a las páginas de inicio de un sitio web, en el segundo nivel el análisis alcanza a todas las páginas que se encuentran un nivel por debajo de la página de inicio de un sitio, y en el tercer nivel nos permite analizar las páginas que se encuentran dos niveles por debajo de la página de inicio de un sitio web.

Settings

<p>Privilege starting points Use privilege starting points and one iteration to find the immediate neighborhood.</p>	<input checked="" type="radio"/> Off (default) <input type="radio"/> On
<p>Perform co-link analysis by Use co-link analysis by page for more specific results. The nodes are likely to be deep pages.</p>	<input type="radio"/> Site <input checked="" type="radio"/> Page (default)
<p>Set iterations Use at least two iterations for locating an issue network.</p>	<input checked="" type="radio"/> 1 (default) <input type="radio"/> 2 <input type="radio"/> 3
<p>Set crawl depth Use a deeper crawl depth for homepages.</p>	<input type="radio"/> 1 <input checked="" type="radio"/> 2 (default) <input type="radio"/> 3

Figura 26. Módulo de configuración de rastreo. Extraído de IssueCrawler.

Además de estas configuraciones el apartado de “*Ceilings*” nos permite definir el número máximo de URLs rastreadas por servidor, el número máximo de URLs rastreadas en total (con un máximo de 60000), el número máximo de de páginas con co-enlaces recuperadas por iteración (con un máximo de 1000) y el número máximo de sitios web con co-enlace por iteración (con un máximo de 1000).

En el apartado “*Crawl Speed*” podemos reducir la prioridad de nuestro análisis en la cola de análisis por realizar de IssueCrawler; y en el apartado “*Exclusion List*” podemos introducir una lista de URLs que serán excluidas del rastreo. Esto se utiliza para eliminar grandes páginas muy enlazadas como buscadores, sitios de descarga de software, etc.

Ceilings (advanced):

Crawled URL ceiling (per host)
The maximum quantity of URLs crawled on each host.

Crawled URL ceiling (per iteration)
The maximum quantity of URLs crawled per iteration.(max 60000)

Co-link ceiling by page (pages per iteration)
The maximum quantity of co-linked pages returned per iteration.(max 1000)

Crawl speed:

Prioritize your crawl **Normal** **Low**

Exclude from network:

Blocks webpages from appearing in your Network

Send crawl completion notification to:

Name crawl:

Launch Crawl

Next step >>
Issue Crawler
Summary

Figura 27. Módulo de configuración avanzada de rastreo. Extraído de IssueCrawler.

IssueCrawler cuenta además con un “Módulo de Gestión de Redes”. Este módulo nos permite almacenar los diferentes análisis que hemos realizado y ver los análisis que hay en cola para ser realizados.

Monday, August 24, 2015 @965.22

Network Manager

Listed: My Networks, most recent at top (10000 total)
Displaying records 1 - 25

List: 25 | 50 | 100 | 200 per page
Next

Network	Date
TransChile_IC_Co-link transparencia.mop.cl (1112), gobiernotransparente.gov.cl (850), mop.cl (719), modernizacion.gob.cl (515), oirs.mop.gov.cl (511)	23 Aug 2015
TransparenciaChile1 interior.gob.cl (36678), frilufstsleiren.no (36348), webhosting.redsalud.gov.cl (24371), transparencia.redsalud.gov.cl (19922), schuetzenmuseum.de (18788)	12 Jul 2014

Top | Next

Figura 28. Módulo de gestión de redes. Extraído de IssueCrawler.

En el módulo “Archivo” se almacenan todos los análisis de los usuarios de IssueCrawler, es decir, los resultados de los análisis obtenidos con IssueCrawler son públicos. Esto obviamente puede ser un problema para aquellos investigadores cuyos objetivos están fuertemente ligados a los resultados obtenidos a través de la herramienta y no quieren hacerlos públicos antes de que sus investigaciones aparezcan en publicaciones indexadas.

Este módulo de archivo cuenta con un motor de búsqueda interno para facilitar la búsqueda y recuperación de los resultados de los usuarios de IssueCrawler y ofrece además una herramienta de programación de análisis que nos permite configurar un análisis en una fecha concreta con el objetivo de permitir la monitorización periódica de una red.

El “Módulo de Visualización e Interactividad” permite generar mapas a partir de los resultados del rastreo. Estos mapas se generan utilizando el formato *Scalable Vector Graphic* (SVG) e incluyen información sobre el nombre del análisis, el autor, fechas de inicio y finalización y los datos de configuración del análisis.

Monday, August 24, 2015 @971.20

Archive

Listed: All networks most recent at top
Displaying records 1 - 25

List: 25 | 50 | 100 | 200 per page
Next

Network	Date	Scheduled
expertise scientifique Faléa + Publishing network		
Pharma I + Publishing network		
underground_home_snow + Publishing network		
Gates Grantees Europe - Co-link - Priv + Publishing network		
Gates Grantees Europe - Co-link - 1 it + Publishing network		

Search Archive

Name of network:

Date Range:
01 Jan 2000
26 Aug 2015

(Part of) Host of Organisation:

Include: All networks
 Scheduled

Sort: Date Name

Figura 29. Módulo de archivo. Extraído de IssueCrawler.

Sunday, August 23, 2015 @897.33

Network Details

TransChile_IC_Co-link [Delete](#)

Author name: Miguel Castillejo
Author email: mc@miguelcastillejo.com
Date: 23 Aug 2015

Co-link Crawl Settings

Co-link analysis by: (page)
Number of iterations: (2)
Crawl depth: (2)
Privileged starting points:(off)

Select Network Depiction

cluster **Cluster Map**
Plot issuecrawl result as spring map.
(1-3 min.)

circle

geo Advanced options

Scheduler

Name of schedule:
TransChile_IC_Co-link

Date Range:
23 Aug 2015
dd mmm yyyy

Every:
 day(s)

Start from:
original starting points

Figura 30. Módulo de visualización e interactividad. Extraído de IssueCrawler.

De los datos que nos proporciona el programa podemos extraer diferentes tipos de resultados. IssueCrawler utiliza el formato .xml para generar los mapas aunque en la última actualización de marzo de 2015 también se ha incluido el formato .gefx

desarrollado por el proyecto Gephi¹⁷. Además del formato .xml también es posible exportar los datos en crudo separados por comas (*raw data*). De estos datos podemos extraer diferentes listas o rankings: ranking de actores en función del número de enlaces entrantes, ranking de actores en función del número de enlaces recíprocos, lista de actores identificados, lista de sitios alcanzados o lista de páginas web recuperadas.

A la hora de generar los mapas también es posible configurar algunos parámetros como el número de nodos que se mostrarán en el mapa, la selección de enlaces en función de su fuerza o de su frecuencia y el tamaño de los enlaces en función del número de enlaces entrantes o enlaces salientes.

Map Editor - Cluster maps (default)

Title:

Author name:

Author email:

Network Data

53	Nodes in the network
100	Node pages
268	Linkages within the network

Choose the quantity of nodes to be mapped according to a significance measure based on inlink count per node.

Choose nodes to be mapped

total network (default)

top 20 nodes

top 50 nodes

top 100 nodes

top 200 nodes

Selection of ties by specificity is the qualitative strength of ties. The network clusters actors with strongest ties to one another.

Selection of ties by specificity

total set

select top 50%

select top 30%

select top 15%

Selection of ties by frequency is the quantitative force of ties. The network clusters actors with the greatest quantity of ties between them.

Selection of ties by frequency

total

frequency > 1

frequency > 3

frequency > 10

Node size is relative to quantity of links received, or to the links given and received.

Size of nodes by

Inlinks

Inlinks and Outlinks

Figura 31. Opciones de configuración del módulo de visualización e interactividad. Extraído de *IssueCrawler*.

Cada uno de los nodos del mapa generado es interactivo y si hacemos clic sobre él se despliega un pop-up con la dirección URL asociada con el nombre del nodo, el número de enlaces entrantes y los enlaces que hace y recibe entre el resto de actores de la

¹⁷ Gephi es una plataforma interactiva de visualización y exploración de redes, sistemas complejos y gráficos dinámicos y jerárquicos. Disponible en <http://gephi.github.io/>

red. Es posible activar y desactivar los nodos en función de que estos correspondan a dominios o subdominios de un sitio y hacer zoom en diferentes partes del mapa. Los mapas generados pueden ser almacenados en el “Módulo de Gestión de Redes y Archivo”, ser imprimidos o exportarse en diferentes formatos: .svg, .jpg, .png, .tiff, .pdf, .html. o .net

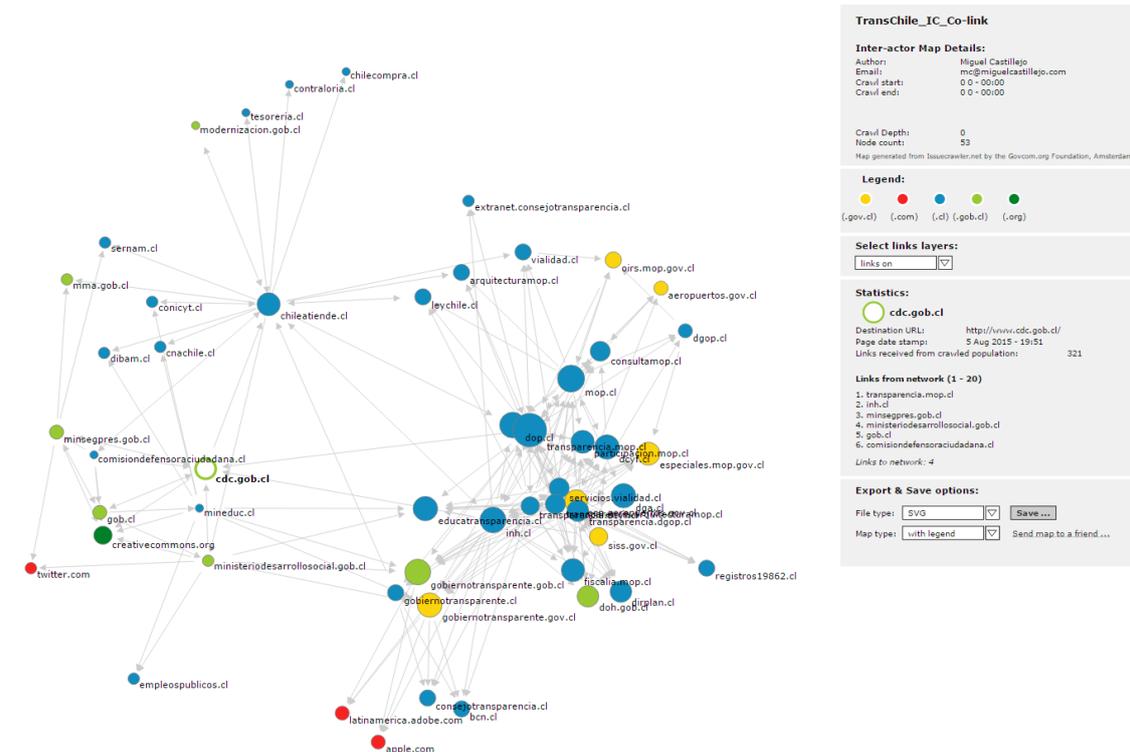


Figura 32. Ejemplo de visualización redes generado con IssueCrawler.

b) SocSciBot

SocSciBot es un crawler diseñado por el Statistical Cybermetrics Research Group de la Universidad de Wolverhampton en el Reino Unido. Este grupo del ámbito de las Ciencias de la Información está dirigido por Mike Thelwall y busca desarrollar herramientas y metodologías para el aprovechamiento de la información contenida en la Web en la investigación en Ciencias Sociales. Sus principales líneas de investigación son la búsqueda de métricas alternativas, la cienciometría, el análisis de enlaces, la cibermetría, la webometría, el estudio de imágenes de medios sociales y el análisis de sentimiento.

SocSciBot es un software de código cerrado pero gratuito para usos no comerciales. Es posible descargarlo a través del sitio web del Statistics Cybermetrics Research Group en un archivo ejecutable previa solicitud y entrega de datos personales a través de un formulario. La utilización de SocSciBot implica la aceptación de una serie de compromisos éticos y para ello SocSciBot envía un email a los sitios que se analizan con la herramienta entregando la información personal que has aportado al descargar la herramienta para que los sitios rastreados puedan dirigirse a ti en caso de que no estén de acuerdo con el análisis que realizas. Además de esto el Statistical Cybermetrics Research Group monitorea el uso que se hace de SocSciBot aunque se compromete a no revelar esta información a terceros excepto en caso de que se produzcan quejas por un uso no ético de la herramienta. En caso de que esto se produzca la copia de SocSciBot puede ser remotamente deshabilitada de tu ordenador sin aviso previo. A pesar de todas estas precauciones el Statistical Cybermetrics Research Group afirma no haber hecho efectiva nunca esta posibilidad.

SocSciBot funciona únicamente sobre sistemas operativos Windows y puede efectuar un análisis de un máximo de 15.000 páginas con un límite de tiempo de recuperación por página. Esto significa que si una página no puede ser recuperada en un tiempo determinado, por ejemplo cinco segundos, la herramienta pasará a recuperar la siguiente página de la lista.

El objetivo de esta limitación está relacionado con los compromisos éticos de SocSciBot que busca no sobrecargar los servidores que almacenan las páginas que se analizan. Otros compromisos éticos de SocSciBot son cumplir los protocolos "robots.txt" y "HTML robots metatags" que implican no rastrear aquellas partes de un sitio web que el propietario no quiere que sean rastreadas y no recuperar las páginas que el autor ha indicado que no deben ser recuperadas. Además de esto SocSciBot incluye la dirección de correo del autor del rastreo en las peticiones de recuperación de una página, lo que permite que el propietario del sitio pueda contactar al autor del rastreo. En busca de los mismos compromisos éticos, al comienzo de un rastreo SocSciBot envía un aviso al sitio web indicando que el sitio va a ser rastreado lo que permite propietario del sitio cancelar el rastreo. Finalmente, el rastreo que lleva a cabo SocSciBot tiene limitada su velocidad a 86400 páginas por día, una por segundo, con el fin de no sobrecargar los servidores que almacenan los sitios rastreados.

SocSciBot se compone de un módulo de rastreo, un módulo de análisis de enlaces y un módulo de análisis de textos. El análisis que lleva a cabo se compone de dos fases, una primera de rastreo en la que podemos analizar un único sitio web o varios a través de una lista de semillas; y una segunda fase de análisis que puede estudiar bien los enlaces de los sitios rastreados o los contenidos de las páginas recuperadas a través del módulo de análisis de contenido Cyclist.

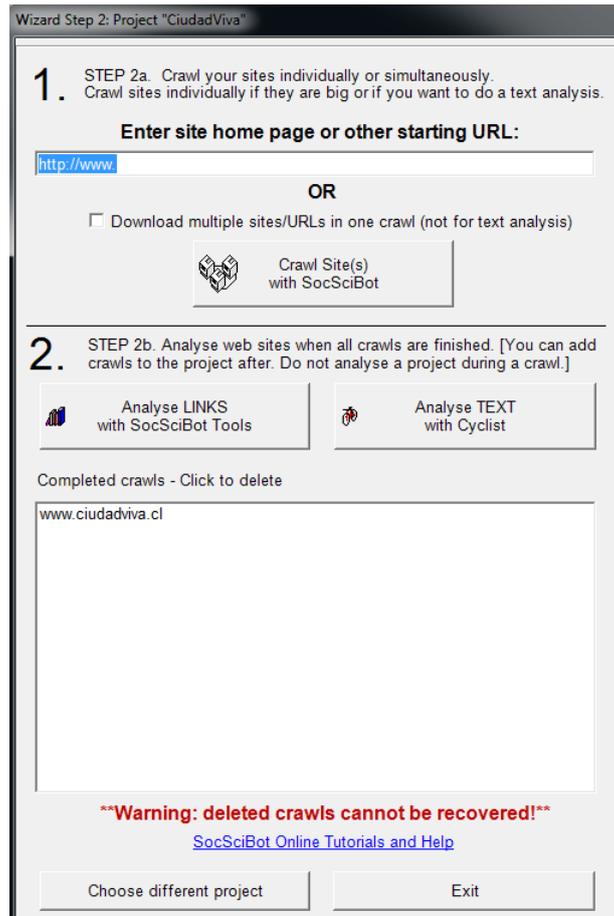


Figura 33. Módulo de rastreo. Extraído de SocSciBot.

Como señalábamos anteriormente, el módulo de “Rastreo” nos permite analizar un único sitio web o varios sitios web. Para llevar a cabo un análisis de varios sitios debemos crear una lista de los sitios que queremos analizar en un archivo de texto con una URL por línea, el cual debe introducirse como input en el programa. En este módulo es posible configurar el rastreo para ignorar determinadas URLs que igualmente deberemos introducir como input en un archivo de texto utilizando la opción “*banned list*”. Elegir el número máximo de URLs a analizar o la profundidad del rastreo

dentro de un sitio web: 0 para la páginas iniciales únicamente, 1 para las páginas que se enlazan con la página inicial, 2 para las que se enlazan con estas últimas y así sucesivamente.

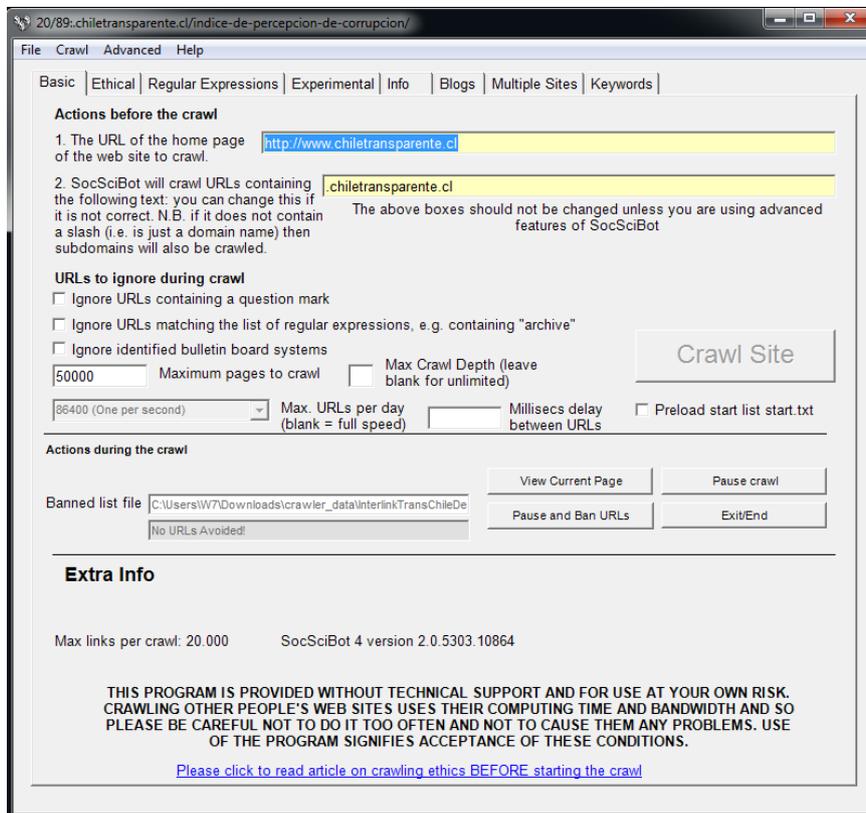


Figura 34. Opciones de configuración del rastreo. Extraído de SocSciBot.

Además de esta opción, el programa cuenta con una lista de expresiones regulares que en caso de ser activada filtrará de los resultados del rastreo aquellas páginas que contengan en su URL alguna de las expresiones de la lista o bien las URLs que correspondan con archivos de los tipos que vemos en la Figura 35.

En el módulo de "Análisis de Enlaces" debemos en primer lugar elegir el tipo de enlaces que queremos considerar en el análisis: enlaces dentro de un sitio web (site self links), enlaces entre los sitios introducidos como semillas (inter-site links of sites crawled) o enlaces entre los sitios pertenecientes al grupo de semillas y sitios no pertenecientes al grupo de semillas (inter-site links of sites not crawled). Además de estas opciones podemos desestimar los enlaces que se dirigen a las páginas de inicio de un sitio web y elegir la forma en que agrupamos los resultados de las URLs: por página, por directorio, por dominio o por sitio.

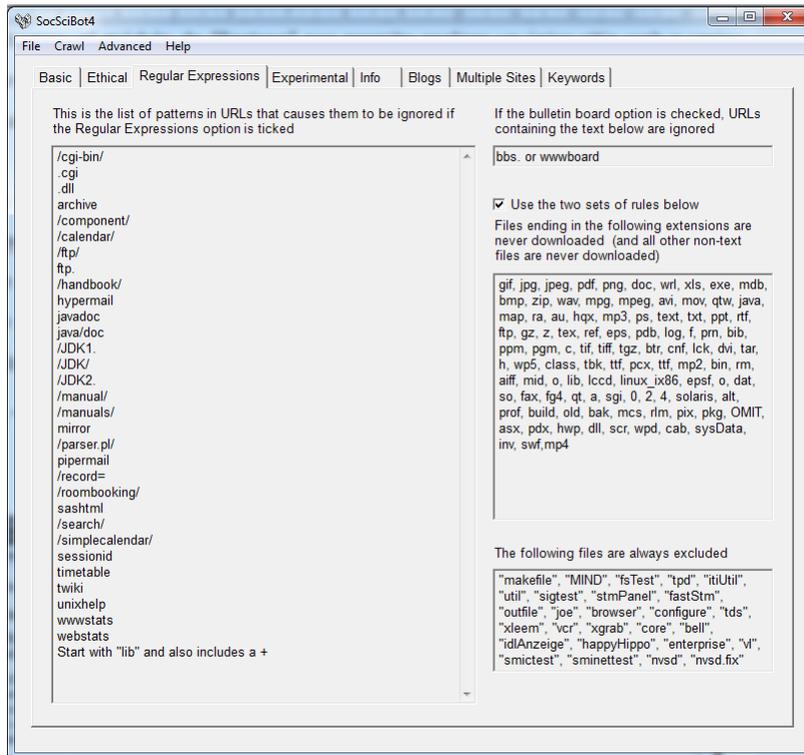


Figura 35. Lista de expresiones regulares. Extraído de SocSciBot.

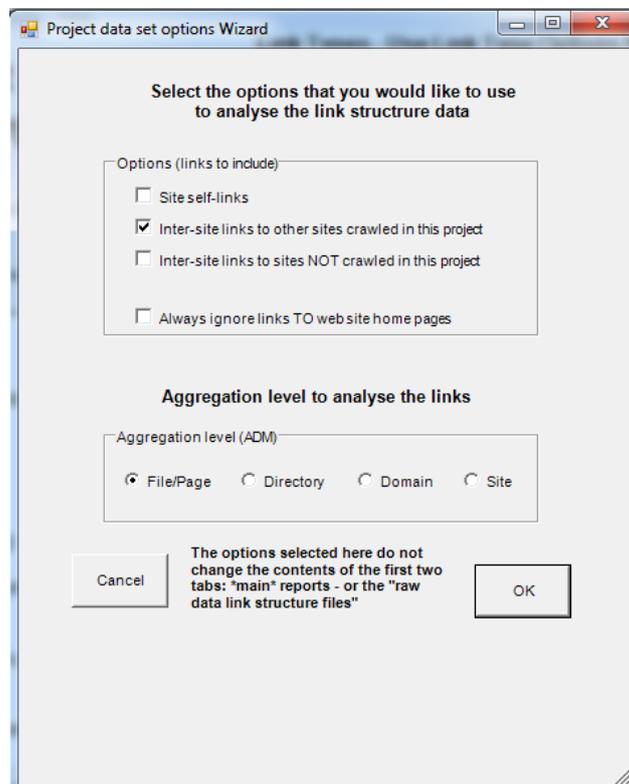


Figura 36. Módulo de análisis de enlaces. Extraído de SocSciBot.

En la primera pestaña del módulo de “Análisis de Enlaces” podemos ver los principales informes que ofrece SocSciBot:

- **Page and link counts:** ofrece el número de páginas y de enlaces salientes en cada sitio.
- **ADM count summary:** ofrece el número de enlaces entrantes y salientes de cada sitio utilizando los cuatro tipos de agrupamiento de URLs que señalábamos (página, directorio, dominio o sitio).
- **All external links:** ofrece una lista de los enlaces externos de cada sitio agrupados por las cuatro opciones ya señaladas y los enlaces salientes dirigidos a sitios no pertenecientes al grupo de sitios rastreados.
- **Directory interlinking:** ofrece el número de enlaces entre parejas de directorios en sitios diferentes.
- **Domain interlinking:** ofrece el número de enlaces entre parejas de dominios en sitios diferentes.
- **File interlinking:** ofrece el número de enlaces entre parejas de páginas web en sitios diferentes
- **Site interlinking:** ofrece el número de enlaces entre parejas de sitios.
- **Selected external links:** ofrece una lista de los enlaces salientes en función del tipo de agrupamiento elegido en la fase inicial del análisis de enlaces, recordamos: página, directorio, dominio o sitio.
- **Selected external links with counts:** ofrece una lista de los enlaces salientes y la frecuencia de cada uno también en función del nivel de agrupamiento.
- **Unselected external links:** ofrece una lista de los enlaces salientes dirigidos a páginas, directorios, dominio o sitios (en función de nuestra elección) no pertenecientes al grupo de sitios rastreados.

- **Unselected external links with counts:** ofrece una lista de los enlaces salientes dirigidos a páginas, directorios, dominio o sitios (en función de nuestra elección) no pertenecientes al grupo de sitios rastreados con las frecuencias de cada uno de ellos.

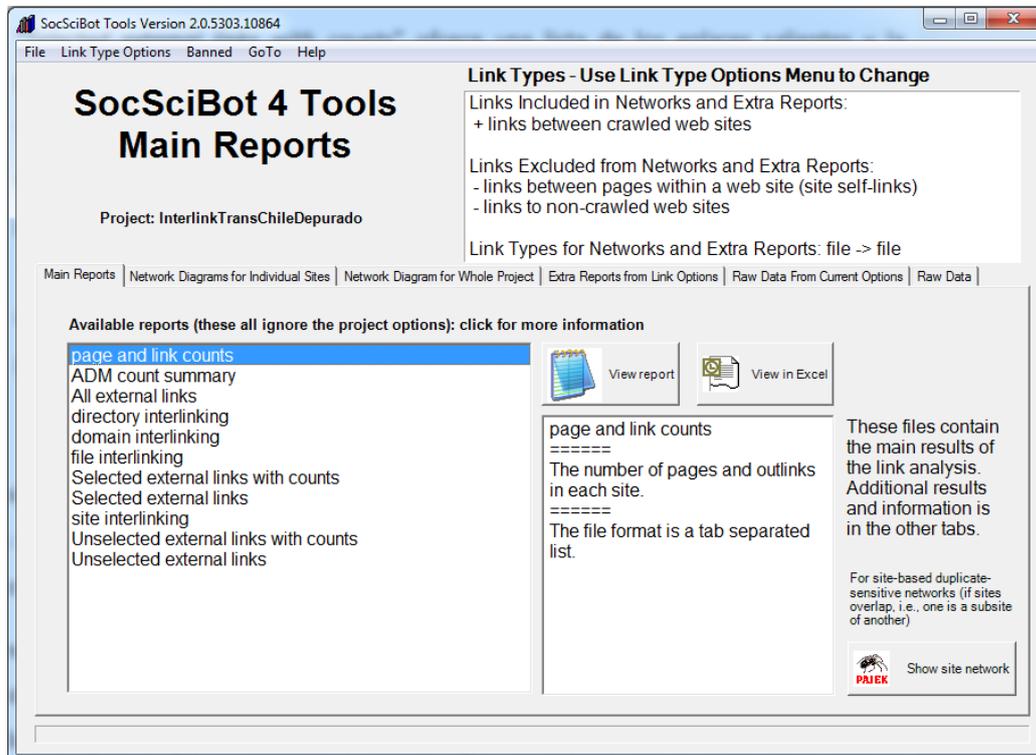


Figura 37. Módulo de informes. Extraído de SocSciBot.

En el mismo módulo podemos visualizar la red de enlaces entre los sitios rastreados, la red de enlaces salientes de los sitios uno a uno o la red de enlaces incluyendo los enlaces dirigidos a sitios no rastreados. Todas estas opciones pueden visualizarse a través del módulo de visualización de SocSciBot o utilizando Pajek¹⁸, UCINET¹⁹ o cualquier otro software que utilice el formato .net.

¹⁸ Pajek un software para el análisis y visualización de redes sociales desarrollado en la Universidad de Ljubljana por Vladimir Batagelj, Andrej Mrvar y Matjaž Zaveršnik. Disponible en <http://mrvar.fdv.uni-lj.si/pajek/>

¹⁹ UCINET es un software para el análisis y visualización de redes sociales desarrollado por Borgatti, Everett y Freeman. Disponible en <https://sites.google.com/site/ucinetsoftware/>

El módulo de “Visualización” que utiliza SocSciBot es el Webometric Analyst Network Graph Tool pero en una versión anterior a la que incorpora Webometric Analyst. Posteriormente analizaremos Webometric Analyst incluido su módulo de visualización así que no nos detendremos a analizar las diferencias entre las dos versiones que consideramos menores.

En el módulo de “Análisis de Contenidos”, también llamado “Cyclist”, podemos indexar los contenidos de las páginas recuperadas en la fase de rastreo para crear un corpus lingüístico. Para ello nos pedirá que introduzcamos como input un archivo de texto con una lista de “palabras vacías” o “stopwords”, estas son palabras que no tienen contenido semántico como preposiciones, conjunciones, artículos, etc. Además de la posibilidad de eliminar las palabras vacías, Cyclist aplica una lematización básica eliminando los plurales y las palabras que contienen números.

Una vez indexado el contenido y creado el corpus es posible realizar búsquedas de palabras a través del motor de búsqueda que incorpora este módulo. El motor de búsqueda nos devolverá las páginas indexadas que contienen la palabra buscada aportando en los resultados la URL en cuestión, el código HTML de la página y las 100 primeras palabras de la página. En este módulo podemos además generar informes con el número total de palabras que componen el corpus, con la frecuencia con que aparece cada palabra, o filtrar todas las palabras que contengan una raíz o un texto específico.

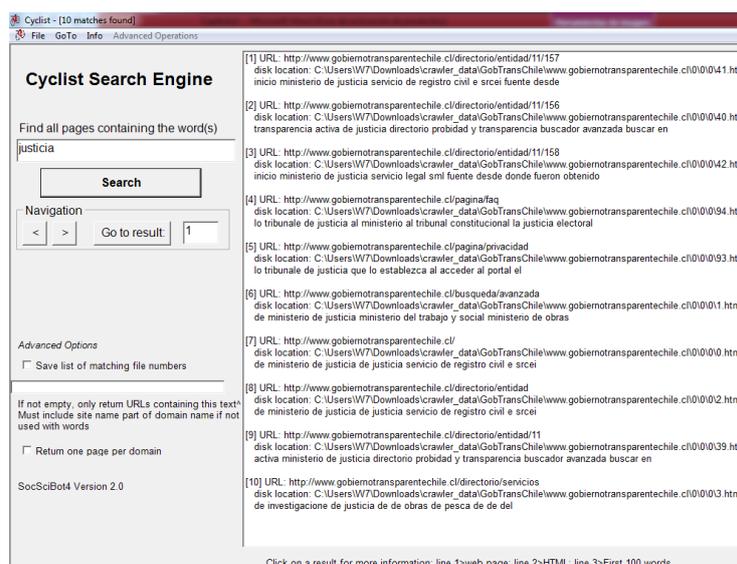


Figura 38. Módulo de análisis de contenido Cyclist. Extraído de SocSciBot.

c) Webometric Analyst

Webometric Analyst es otra de las herramientas desarrolladas por el Statistical Cybermetrics Research Group de la Universidad de Wolverthampton que dirige Mike Thelwall. A diferencia del crawler SocSciBot, esta herramienta se basa en los resultados obtenidos a través del API de búsqueda de Bing (Thelwall, 2012) para ofrecer varios tipos de análisis webmétricos que incluyen plataformas como como Twitter, Youtube o Mendeley. Puesto que en esta investigación nos limitamos a analizar redes temáticas a través del análisis de redes de hiperenlaces entre sitios web, no analizaremos en detalle las características de funcionamiento específicas del análisis de cada una de estas plataformas y nos limitaremos a las funcionalidades de la herramienta que se pueden aprovechar en el tipo de análisis que llevamos a cabo.

Tradicionalmente, los grandes buscadores del mercado ofrecían la posibilidad de realizar búsquedas automáticas de hiperenlaces a través de sus crawlers. Sin embargo, debido a la sobreexplotación de las búsquedas por parte de empresas comerciales, no siempre con objetivos éticos, este servicio fue eliminado primero por Google en 2009 y después por Yahoo durante su transición para integrarse en Bing en 2012.

Ante este hecho, y para ofrecer una alternativa que permita seguir desarrollando investigaciones sobre redes de enlaces en la web, Mike Thelwall ha propuesto la utilización de “URL Citations” y de “Title Mentions” (Thelwall, 2011) que es posible consultar a través del API de búsqueda de Bing²⁰ utilizando la plataforma de venta de datos Microsoft Azure Marketplace²¹. Esta plataforma ofrece entre otros muchos servicios 5.000 consultas al API de búsqueda de Bing gratuitas al mes y una tarifa de precios para un mayor número de búsquedas.

Una cita de URL o “URL citation” es la aparición de una URL en una página web sin la asociación con un hiperenlace. Recordemos que entendemos por hiperenlace la porción del código fuente de una página web (elemento <a> más atributo href) que

²⁰ El API de búsqueda de Bing permite acceder a los datos del buscador a través de consultas personalizadas. Disponible en <https://datamarket.azure.com/dataset/bing/search>

²¹ Microsoft Azure™ Marketplace es una plataforma de compra y venta en línea de aplicaciones y conjuntos de datos. Disponible en <https://datamarket.azure.com/home>

activa un nuevo destino de navegación si el usuario hace clic sobre el mismo. Una página web A puede contener la URL de una página web B sin contener necesariamente un hiperenlace dirigido a ella.

De la misma forma, una Mención de título es la aparición del nombre de una página web A en una página web B sin conservar las características de una URL. Es decir, un sitio web puede contener la mención de otro sitio web como W3C sin añadir el dominio “.com” al nombre del mismo. Una diferencia fundamental es que mientras que las Menciones de título hacen referencia a una entidad las citas de URL hacen referencia al sitio web de una entidad.

Thelwall comprobó parcialmente, porque su investigación analizaba únicamente departamentos de Ciencias de la Información de Estados Unidos y universidades del Reino Unido, que existía una correlación de similitud entre los resultados obtenidos con las Menciones de Título, las Citas de URL e los Hiperenlaces. Estos datos fueron comparados además con un ranking de universidades demostrando que las tres técnicas ofrecían resultados correlativos con otras formas de medición más elaboradas y por supuesto también más costosas como son los sistemas tradicionales de evaluación académica.

Como señalábamos, Webometric Analyst ofrece herramientas y utilidades para llevar a cabo análisis webométricos de sitios web pero también en plataformas sociales como Twitter o Youtube, plataformas académicas como Mendeley, ResearchGate y Academia, servicios de búsqueda especializada como GoogleBooks o ArXiv e incluso un servicio de detección de plagio. Lo vemos en la Figura 39 a continuación.

Centrándonos exclusivamente en el módulo de análisis de sitios web, que como decíamos es el que se relaciona directamente con el análisis de redes temáticas, Webometric Analyst ofrece dos interfaces de uso diferentes: interfaz clásica e interfaz reducida o wizard. Probablemente, la razón de incorporar dos interfaces se debe a la multitud de servicios que ofrece la herramienta y a la dificultad de aprendizaje y uso de alguno de ellos.

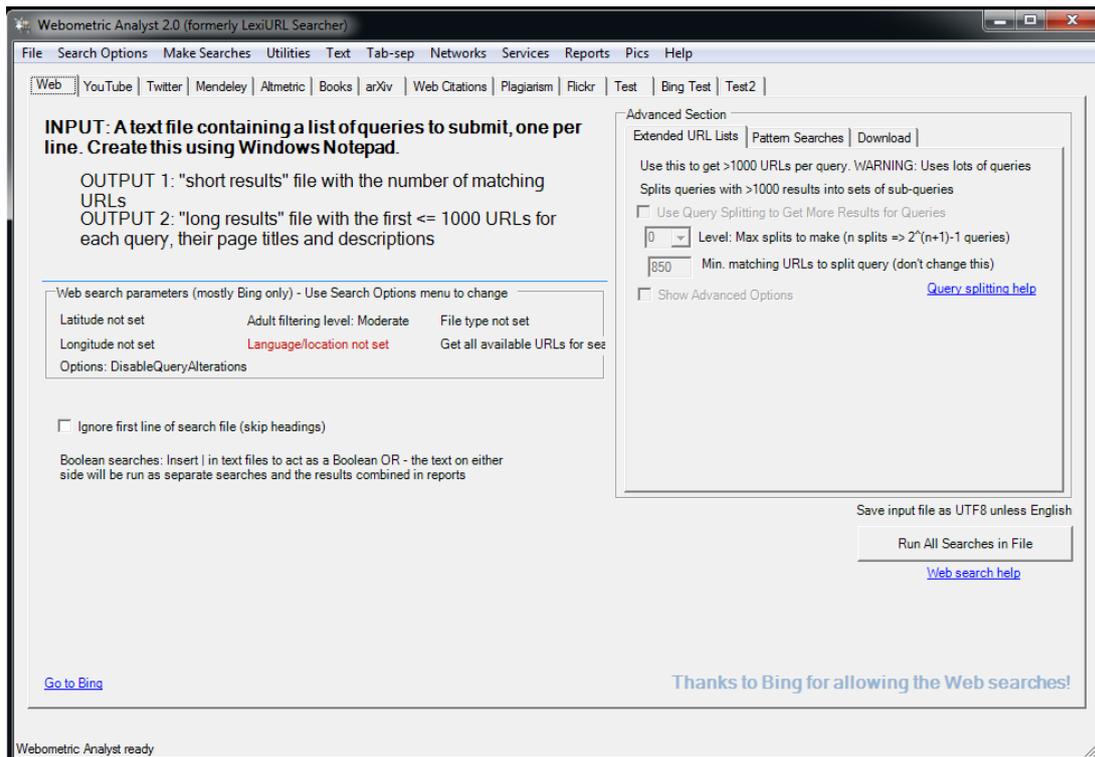


Figura 39. Interfaz avanzada. Extraído de Webometric Analyst.

La interfaz reducida o wizard es la primera ventana que aparece cuando ejecutamos el programa y en ella se ofrecen los principales tipos de análisis de sitios web que se pueden desarrollar con la herramienta dejando fuera los servicios destinados al análisis en las plataformas sociales y servicios de búsqueda que señalábamos anteriormente. Pero para poder seleccionar alguno de los análisis que se muestran, en primer lugar debemos introducir nuestra clave personal de Windows Azure que obtendremos en Microsoft Market Azure después de crear una cuenta o utilizando una cuenta existente de Microsoft. Lo vemos en la Figura 40 a continuación.

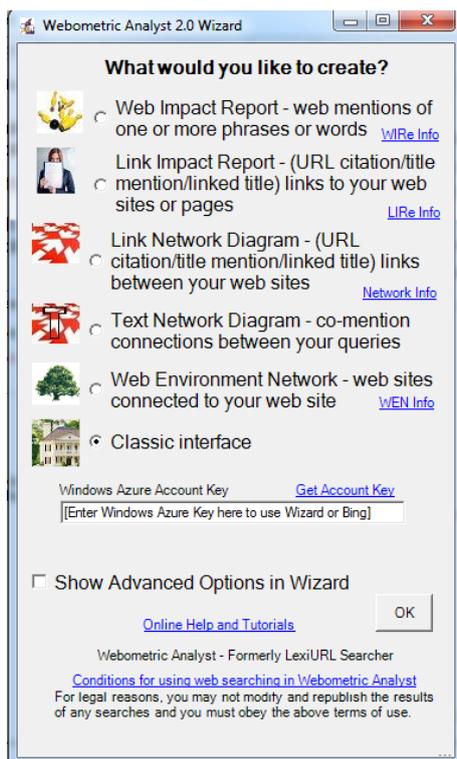


Figura 40. Interfaz reducida wizard. Extraído de Webometric Analyst.

Los análisis que ofrece esta interfaz reducida o wizard son: informes de impacto de palabras clave, informes de impacto de sitios web, diagramas de redes de enlaces, diagramas de redes de co-enlaces y diagramas de enlaces de un sitio web.

Web Impact Report: el informe de impacto de palabras claves o frases se produce a través de la identificación y recuento de las páginas web que mencionan las palabras clave o frases que se introducen en la herramienta como input. Para conseguir esto la herramienta convierte cada palabra clave o frase en una consulta que hace al servicio de búsqueda de Microsoft, Live Search. Los resultados obtenidos se muestran a través de una página web que el programa genera donde se pueden organizar de forma diferenciada en base a URLs, dominios, sitios, dominios por sitio, TLDs²², STLDs²³ y dominios aleatorios.

²² TLD es el acrónimo de Top Level Domains y hace referencia a la selección de dominios de primer nivel en la entrega de resultados del análisis. Este tipo de resultado puede ser útil para discriminar sitios de entidades que desarrollan un determinado tipo de actividad como .edu en el caso de organizaciones educativas o .org para organizaciones no gubernamentales sin fines de lucro. Igualmente puede ser útil para identificar entidades de diferentes países como .es en el caso de sitios españoles o .mx para sitios mexicanos.

Link Impact Report: el informe de impacto de enlaces se construye a través de la identificación y recuento de las páginas web que citan alguna de las URLs (*URL citation*) semillas introducidas como input. Para conseguirlo la herramienta convierte cada una de las URLs introducidas como semillas en una consulta que hace a Bing para recuperar hasta un máximo de 1000 resultados por consulta. Para conseguir recuperar estos 1000 resultados el programa debe repetir la consulta a Bing 20 veces pues el buscador ofrece los resultados en paquetes de 50 URLs. Esto aumenta el consumo de consultas en nuestra cuenta de Microsoft Azure que recordemos está limitada a un máximo de 5000 consultas gratuitas mensuales. Los resultados se visualizan a través de una página web creada por el programa y pueden organizarse en base a URLs, dominios, sitios, dominios por sitio, TLDs, STLDs y dominios aleatorios.

Link Network Diagram: los diagramas de redes de enlaces permiten visualizar gráficamente los enlaces que se producen entre un grupo determinado de sitios o páginas web (semillas) que introducimos en el programa como input. Para construir este gráfico el programa convierte cada una de las semillas (A) en una consulta que hace a Bing limitando la búsqueda a otro de los sitios incluido en el grupo de semillas (B). Es decir, si la URL de A aparece en el sitio B se considera que B está dirigiendo un enlace a A. Realizando todas las combinaciones posibles de búsqueda entre las semillas se consigue construir el diagrama de los enlaces entre sitios, algo que en otras herramientas se denomina "*interlinking*". En función del número de enlaces que hay entre un sitio y otro se definirá la fuerza (*strength*) del enlace entre los dos sitios. Los resultados se ofrecen en forma de un mapa de enlaces a través de la interfaz gráfica de Webometric Analyst que describiremos detalladamente más adelante. En estos mapas cada URL está representada por un nodo y los enlaces entre nodos se representan a través de una flecha. Tanto el tamaño de los nodos como el grosor de las flechas puede variar en función de varios criterios, habitualmente el número de enlaces que hace o recibe un sitio web.

²³ STLD es el acrónimo de Second Top Level Domains y hace referencia a la selección de dominios de segundo o primer nivel en la entrega de resultados del análisis. El análisis a partir de este tipo de dominio puede ser útil en los casos en que se combinan dominios como .edu.ar en el caso de entidades educativas argentinas o .org.uk en el caso de organizaciones no gubernamentales sin fines de lucro del Reino Unido.

Text Network Diagram: esta opción funciona de forma similar a Link Network Diagram pero buscando parejas de URLs, es decir, co-enlaces. Para ello la herramienta convierte las semillas que introducimos como inputs en consultas que realiza por parejas en el buscador Bing limitando la búsqueda al resto de URLs que hay en el grupo de semillas. Es decir, la URL de un sitio A y la URL de un sitio web se convierten en una consulta A + B. De esta forma, si un sitio C contiene las URLs de A y B se considera que existe un co-enlace entre A, B y C. Repitiendo este proceso para todas las posibles combinaciones entre las URLs del grupo de semillas se consigue identificar a los sitios web que presentan un co-enlace con alguna de las parejas posibles del grupo de semillas.

Web Environment Network: este tipo de análisis busca identificar los sitios que se pueden considerar más relevantes para un sitio determinado elegido por el usuario. Se consideran sitios web relevantes aquellos que presentan más enlaces con aquellos sitios que contienen URL *citations* de nuestro sitio. Es decir, aquellos sitios más enlazados con los sitios que dirigen un enlace hacia el nuestro. Para conseguirlo, la herramienta identifica en primer lugar los sitios que citan la URL de nuestro sitio (*inlinks* o *inboundlinks*) a través del buscador Bing. Una vez identificados los enlaces entrantes de nuestro sitio la herramienta recupera las páginas web de esos sitios e identifica los enlaces que dirigen a otros sitios externos al nuestro y selecciona los 50 sitios que reciben más enlaces para crear un diagrama que tiene en cuenta las parejas de sitios que reciben enlaces desde un mismo sitio. Es decir, dibuja la red en base a los co-enlaces que existen entre esos 50 sitios seleccionados.

Además de estos análisis principales, Webometric Analyst cuenta con un amplio abanico de utilidades que podríamos caracterizar como instrumentales pues están dirigidas al preprocesamiento y depuración de datos necesarios para el desarrollo de otras fases de análisis. Como ya hemos dicho nuestro objetivo no es describir con detalle todas las funcionalidades de esta herramienta, sino describir aquellas que pueden resultar útiles para el análisis de redes temáticas. A continuación, en la Figura 41 vemos las numerosas utilidades que ofrece Webometric Analyst.

Como hemos visto, algunos de los análisis que hemos descrito previamente no ofrecen una representación gráfica de los resultados en forma de diagrama o mapa de red. Esto se debe a que los resultados que se obtienen necesitan un proceso previo de

depuración para que la representación de los mismos pueda ofrecer la información que se precisa.

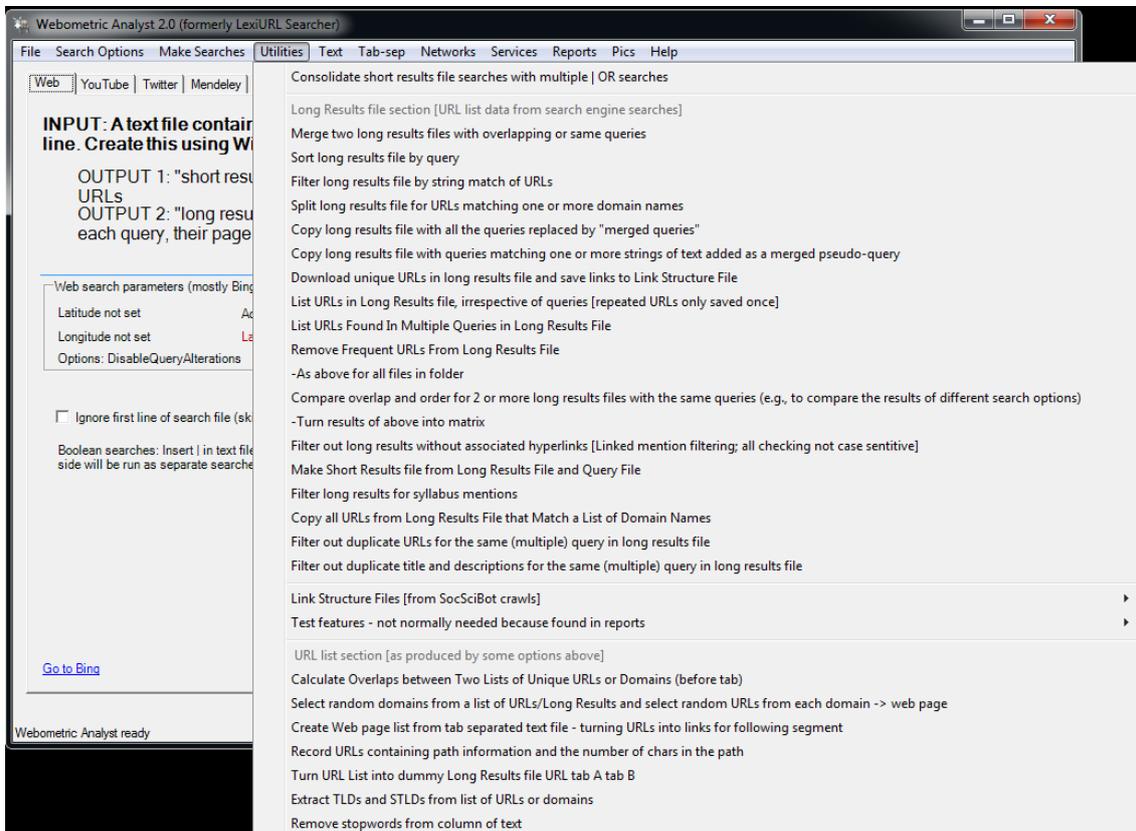


Figura 41. Lista de utilidades. Extraído de Webometric Analyst.

Es el caso de *Link Impact Report*, que recordemos identificaba los enlaces entrantes de uno o varios sitios web. Este tipo de análisis puede ser de gran utilidad para identificar los flujos de información entre organizaciones y los vínculos que se producen entre ellas. Sin embargo, los resultados que se obtienen se ven contaminados por servicios de indexación e información sobre sitios web que indizan y por tanto “citan” una gran cantidad de sitios web sin estar por ello relacionados temáticamente con esos sitios.

Para conseguir representar gráficamente la red de enlaces entrantes de uno o varios sitios web deberemos en primer lugar depurar la lista de URLs que citan los sitios web introducidos como semillas. Para ello podemos hacer uso de la utilidad *List URLs in long list results*. Esta opción extraerá todas las URLs del archivo de texto que el programa nos ofrece como resultado, en el que también aparece la información sobre

la consulta realizada y una parte del texto asociado con cada página. Al extraer las URLs el programa eliminará además todas las URLs repetidas dejando exclusivamente las URLs únicas.

Una vez extraídas las URLs únicas será necesario eliminar aquellas URLs que pertenecen a los sitios de indexación o análisis de dominios que como señalábamos previamente no tienen una relación directa con la semillas a pesar de contener una cita de los mismos. Para ello existe la opción de filtrar los resultados eliminando las URLs que coincidan con una lista que podríamos llamar "*banned URLs*" o "URLs prohibidas". Este filtro que en el programa tiene el nombre de "*List all URLs in file matching specified patterns*" se puede aplicar utilizando URLs, frases o palabras clave.

Una vez depurados los datos extraídos del análisis inicial ya estamos en condiciones de representar gráficamente nuestra red. Para ello podremos utilizar la opción *Convert title or URL citation long results or covariants into Pajek network* que construirá a partir de un archivo de texto un diagrama de red de enlaces entrantes de los sitios web elegidos como semillas.

En el módulo de visualización encontramos un amplio menú con numerosas opciones de configuración. Las más relevantes para el análisis de redes temáticas son las que permiten modificar los nodos, los enlaces o el algoritmo de representación. En relación a las opciones de configuración de los nodos podemos modificar su tamaño en función del número de enlaces entrantes o salientes, modificar el color en función del tipo de dominio o de variables personalizables como el nombre o la URL del nodo o agruparlos en caso de que dos nodos representen diferentes sitios de un único actor de la red. En relación a las opciones de configuración de los enlaces estos pueden ser modificados para ampliar su grosor en función del número de enlaces que existen entre dos nodos. En relación a los diferentes algoritmos de representación de redes, en el menú "*Layout*" encontramos la opción de utilizar el algoritmo Fruchterman Reingold y el algoritmo Kamada Kawai, además de otras opciones de representación que nos permiten visualizar la red en un formato circular o en un formato aleatorio. En el menú "*File*" de este módulo encontramos las opciones de exportación de los gráficos de redes en formato .net, .jpg y .txt.

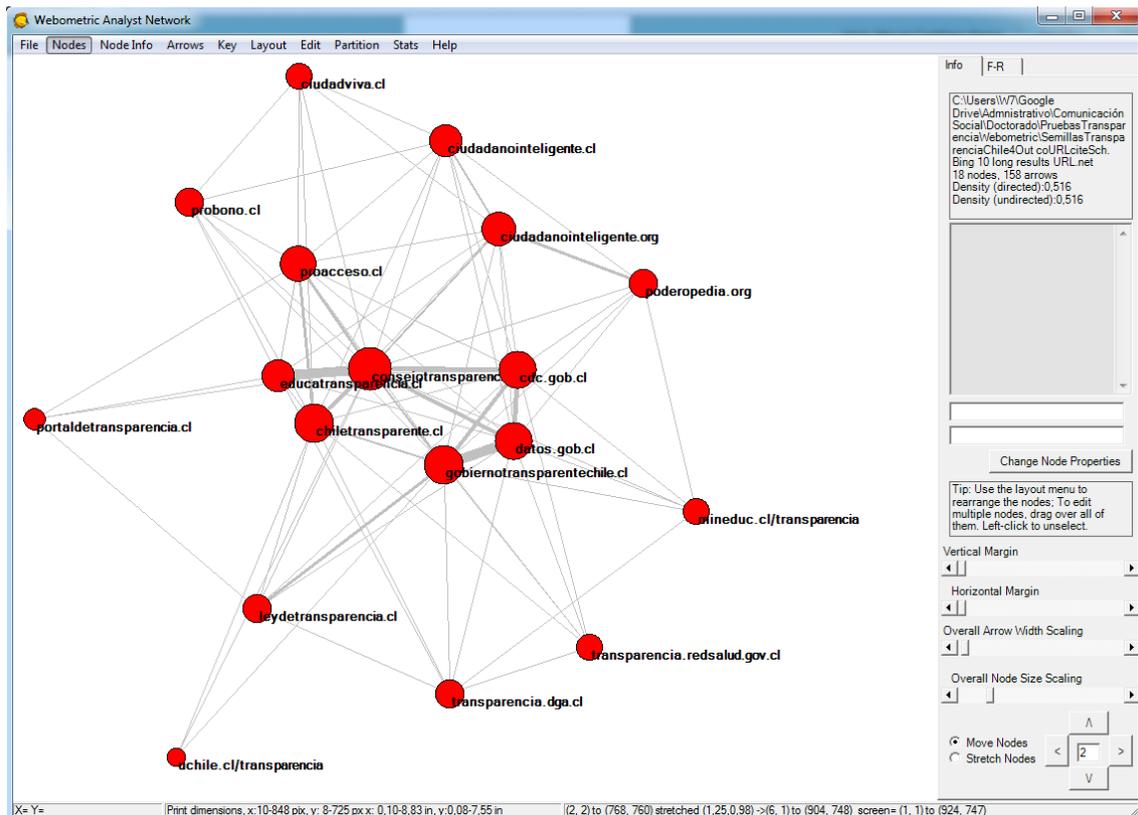


Figura 42. Ejemplo de visualización de redes generado con Webometric Analyst.

Dentro del módulo de visualización encontramos el menú “Stats” que podríamos considerar un módulo diferenciado de análisis estadístico de redes sociales (SNA). En este menú encontramos opciones para calcular diferentes medidas tanto de la red en su conjunto como de los nodos que conforman la red. Entre estas medidas están el cálculo de la centralidad de los nodos, la densidad de la red o el stress de la red si utilizamos el algoritmo Fruchterman-Reingold.

Entre otras opciones que podrían resultar útiles para el análisis de redes temáticas en la web están la de comparar redes (*compare networks*), fusionar redes (*merge networks*) o dividir archivos (*split files*). Las posibilidades de Webometric Analyst son enormes y la variedad de utilidades que incorpora hace de ella una herramienta útil por sí misma o en combinación con otras herramientas. Sin embargo, esta enorme variedad de posibilidades también trae aparejada que el proceso de aprendizaje sea largo y no siempre fructífero.

d) VOSON

Voson System es un crawler de investigación desarrollado dentro del proyecto Visual Observatory for the Study of Online Networks que dirige Roberty Ackland y que pertenece al Australian Demographic and Social Research Institute de la Australian National University. VOSON comenzó su desarrollo en 2005 y fue abierto al público a partir de 2006. Actualmente ofrece dos versiones de la herramienta: como plugin para ser utilizado en conjunción con el software de administración y visualización de redes NodeXL²⁴ y una segunda versión como sistema web en la nube a través de la plataforma web de la empresa Uberlink²⁵.

VOSON requiere de un registro previo tanto para la versión web como para la versión combinada con NodeXL y cuenta con tres tipos de suscripción: *Free*, *Professional* o *Premium*. Los tres tipos de suscripciones ofrecen funcionalidades similares pero cada una de ellas limita el número y tamaño de los análisis que llevamos a cabo a través de lo que denominan *VOSON Activity Units* (VAU).

Estos VAU se miden en función de los segundos de procesamiento y la memoria del sistema requerida para el análisis. Un VAU puede ofrecer 10 segundos de procesamiento consumiendo 10Mb de memoria o 100 segundos de procesamiento consumiendo 1Mb de memoria. La suscripción Free ofrece 50 VAUs gratuitos mensuales, la suscripción Professional ofrece 500 VAUs mensuales por un precio de 20\$ y la suscripción Premium ofrece 5000 VAUs mensuales por un precio de 99\$. En caso de que agotemos nuestros VAUs también es posible comprarlos sin cambiar el tipo de suscripción.

VOSON utilizaba los datos ofrecidos por el API de Yahoo para desarrollar los análisis de hiperenlaces pero como ya dijimos este servicio fue retirado tras la integración de Yahoo en Bing. Actualmente VOSON utiliza los datos del crawler de la empresa de

²⁴ NodeXL es un software gratuito y de código abierto que funciona en combinación con Excel a través de una plantilla y permite la administración y visualización gráficos de redes. <https://nodexl.codeplex.com/>

²⁵ Uberlink Corporation es una empresa creada en Australian National University que ofrece herramientas y servicios de consultoría para el análisis empírico de redes sociales y organizacionales. <http://www.uberlink.com/>

marketing especializada en SEO Moz que ofrece como servicio a través de su API Mozscape²⁶.

Las dos versiones de VOSON que señalábamos inicialmente, la versión web y la versión combinada con NodeXL, ofrecen funcionalidades muy similares y por ello vamos a caracterizar únicamente la versión web de la herramienta que consideramos más adecuada desde la perspectiva del investigador por su mayor usabilidad y su más fácil proceso de instalación.

VOSON incorpora un módulo de minería web y extracción de datos configurable en numerosos parámetros que permite la búsqueda y recuperación de hiperenlaces y el análisis de contenido de las páginas recuperadas, un módulo de visualización gráfica de redes con herramientas de edición y exportación y un módulo de análisis de redes sociales (SNA) que ofrece diferentes medidas relacionadas con las características de la red.

El módulo de minería web funciona a través de la creación de bases de datos que se almacenan en nuestra cuenta de VOSON. Para comenzar el análisis debemos crear una base de datos y configurar las características del rastreo que desarrollará la herramienta. VOSON, a diferencia del resto de crawlers que hemos analizado, permite desarrollar análisis de enlaces entrantes y salientes en un mismo análisis. Para ello deberemos seleccionar la opción "*crawl inbound*" y la opción "*crawl outbound*".

Las opciones de configuración del análisis de enlaces entrante "*crawl inbound*" nos permiten limitar el número de enlaces que se van a recuperar con un máximo de 2500 enlaces entrantes por página y nos permite definir si queremos analizar las páginas contenidas en un sitio web o únicamente la página de inicio.

En las opciones de configuración del rastreo de enlaces salientes "*crawl outbound*", podemos limitar el número de enlaces salientes de cada página que serán analizados hasta un máximo de 1000 páginas, el número máximo de páginas no productivas "*unproductive pages*", esto es el número de páginas que no ofrecen respuesta a la

²⁶ Mozscape es la API que permite extraer datos del crawler propiedad de la empresa de marketing Moz. <https://moz.com/products/api>

consulta del crawler, el número máximo de páginas que serán rastreadas en un sitio web "*depth crawl pages*" hasta un máximo de 1000 páginas, el nivel de profundidad del rastreo "*depth crawl levels*", esto es si queremos analizar únicamente las páginas iniciales de un sitio (nivel 1), las páginas internas de un sitio enlazadas con la página de inicio (nivel 2) o las páginas internas de un sitio que enlazan con las páginas de segundo nivel (nivel 3) hasta un nivel máximo de profundidad de cuatro niveles. Es posible además restringir el rastreo al nivel de subdirectorio en caso de que queramos introducir el subdirectorio de un sitio web como semilla, o analizar todas las páginas internas de un sitio web evitando las limitaciones de rastreo que nos ofrecían las posibilidades de configuración anteriores.

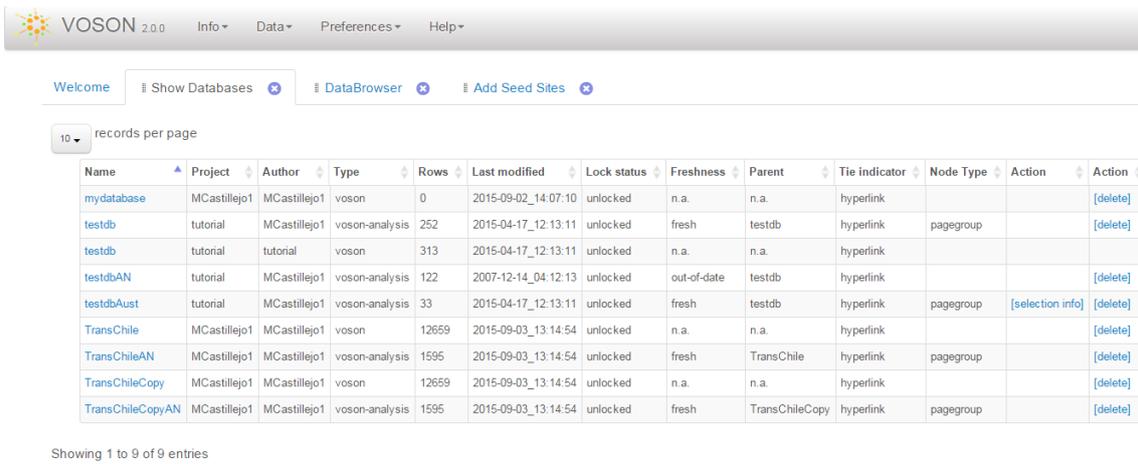
Otra posibilidad que nos ofrece la herramienta en este módulo es la de crear una red teniendo en cuenta únicamente las semillas o crear una red considerando únicamente los sitios que enlazan con dos o más de los sitios del grupo de semillas. Esta opción (*automatically create seeds and seeds + important networks*) puede ser utilizada para llevar a cabo análisis de coenlaces salientes y coenlaces entrantes para lo que simplemente tendremos que seleccionar uno de los dos tipos de rastreo de enlaces entrantes o salientes y considerar la red formada a partir de esos sitios que enlazan con al menos dos de los sitios del grupo de semillas.

Una vez creada la base de datos y configuradas todas las opciones del rastreo deberemos introducir una selección de URLs que utilizaremos como semillas para lo que debemos utilizar la opción "*Add seed sites*". En esta interfaz debemos introducir las URLs seleccionadas y se nos ofrece la posibilidad de modificar nuevamente las opciones del rastreo. Una vez introducidas las semillas y confirmadas las opciones de configuración debemos activar la opción "*Ready to crawl*" y nuestro análisis pasará a la cola de espera de VOSON. Cuando el rastreo se haya completado recibiremos un aviso a través de la cuenta de correo electrónico que hayamos utilizado para registrarnos en VOSON. El tiempo de espera para la obtención de los resultados dependerá del tamaño del análisis pero también del tipo de suscripción que tengamos. Los análisis realizados con una cuenta "*Free*" tienen una prioridad baja, los análisis realizados con una cuenta "*Professional*" tienen una prioridad media y los análisis realizados con una cuenta "*Premium*" tienen prioridad alta.

Figura 43. Opciones de configuración de rastreo. Extraído de VOSON.

Figura 44. Módulo de selección de semillas. Extraído de VOSON.

Una vez se ha realizado el rastreo y hemos recibido la notificación en nuestra cuenta de correo veremos que en la interfaz de visualización de bases de datos se ha creado una nueva base de datos junto a la creada por nosotros inicialmente. VOSON trabaja con dos tipos de bases de datos, una que contiene el total de los resultados obtenidos y cuyo atributo “*Type*” es igual a “*voson*”; y un segundo tipo con los resultados depurados del análisis cuyo atributo “*Type*” es “*voson-analysis*”. El nombre de esta segunda base de datos tomará el nombre de la base de datos original más el sufijo “AN” diminutivo de *analysis*. Los vemos a continuación en la Figura 45.



VOSON 2.0.0 Info Data Preferences Help

Welcome Show Databases DataBrowser Add Seed Sites

10 records per page

Name	Project	Author	Type	Rows	Last modified	Lock status	Freshness	Parent	Tie indicator	Node Type	Action	Action
mydatabase	MCastillejo1	MCastillejo1	voson	0	2015-09-02_14:07:10	unlocked	n.a.	n.a.	hyperlink			[delete]
testdb	tutorial	MCastillejo1	voson-analysis	252	2015-04-17_12:13:11	unlocked	fresh	testdb	hyperlink	pagegroup		[delete]
testdb	tutorial	tutorial	voson	313	2015-04-17_12:13:11	unlocked	n.a.	n.a.	hyperlink			
testdbAN	tutorial	MCastillejo1	voson-analysis	122	2007-12-14_04:12:13	unlocked	out-of-date	testdb	hyperlink			[delete]
testdbAust	tutorial	MCastillejo1	voson-analysis	33	2015-04-17_12:13:11	unlocked	fresh	testdb	hyperlink	pagegroup	[selection info]	[delete]
TransChile	MCastillejo1	MCastillejo1	voson	12659	2015-09-03_13:14:54	unlocked	n.a.	n.a.	hyperlink			[delete]
TransChileAN	MCastillejo1	MCastillejo1	voson-analysis	1595	2015-09-03_13:14:54	unlocked	fresh	TransChile	hyperlink	pagegroup		[delete]
TransChileCopy	MCastillejo1	MCastillejo1	voson	12659	2015-09-03_13:14:54	unlocked	n.a.	n.a.	hyperlink			[delete]
TransChileCopyAN	MCastillejo1	MCastillejo1	voson-analysis	1595	2015-09-03_13:14:54	unlocked	fresh	TransChileCopy	hyperlink	pagegroup		[delete]

Showing 1 to 9 of 9 entries

Figura 45. Módulo de visualización de bases de datos. Extraído de VOSON.

VOSON crea esta base de datos de análisis utilizando unos parámetros de configuración predeterminados pero podemos crear nuevas bases de datos de análisis modificando estos parámetros a partir de la base de datos original que contiene la totalidad de los resultados. Para hacerlo debemos utilizar la opción “*data/create/voson-analysis*” donde podremos además decidir si queremos que VOSON agrupe las páginas que comparten dominio o preferimos obtener resultados de ellas por separado. En la opción “*data/create/voson-analysis*” aparecen otras posibilidades de configuración, enlaces recíprocos entre sitios, análisis de coenlaces y análisis de marcos conceptuales “*frame component*”, que a la fecha de esta investigación no ofrecen resultados pertinentes y se encuentran en proceso de desarrollo según confirmó el director del proyecto Richard Rogers a una consulta hecha a través de las vías de contacto sugeridas en la documentación del software.

Al seleccionar y abrir una de las bases de datos veremos varias informaciones relacionadas con las páginas rastreadas por la herramienta. Entre estas informaciones están el tipo de dominio, la iteración de análisis en que se recuperó²⁷, el grupo de páginas al que pertenece, el número de enlaces entrantes, el número de enlaces salientes, el título de la página, las palabras claves de la página, la descripción de la página, las palabras que contiene con sus frecuencias y el número de páginas que

²⁷ VOSON desarrolla su rastreo de enlaces en dos iteraciones analizando en la primera iteración los enlaces de las páginas que se introducen como semillas y analizando en la segunda iteración los enlaces de las páginas que enlazan con las semillas. Esto produce tres “anillos” de resultados en los que las URLs del grupo de semillas forman el “Anillo 1”, las páginas que enlazan con las semillas forman parte del “Anillo 2” y las páginas que enlazan con este último grupo forman el Anillo 3.

forman parte del grupo de páginas al que pertenece. Todas estas informaciones se muestran en un formato típico de bases de datos con filas y atributos como vemos a continuación en la Figura 46.

Row	ID	URL	Pagegroup	PagegroupID	ccTLD code	genericTLD code	Crawl Status	Ringset	Pagegroup Description	Indegree	Outdegree	Title	Page Keywords	Page Description
1	1	http://www.ciudadanointel...	http://www.cdc.gob.cl/	1	undefined	org	1	1	X	836	73	Fundación Ciudadano Inle...		Usamos tecnolog para l...
2	2	http://www.ciudadanointel...	http://www.cdc.gob.cl/	1	Chile	undefined	1	1	X	188	73	Fundación Ciudadano Inle...		Usamos tecnolog para l...
3	3	http://www.gobiernotransp...	http://www.cdc.gob.cl/	1	Chile	undefined	1	1	X	1067	95	Gobierno de Chile - Direc...		
4	4	http://www.consejotranspa...	http://www.cdc.gob.cl/	1	Chile	undefined	1	1	X	994	0			
5	5	http://www.chiletranspare...	http://www.cdc.gob.cl/	1	Chile	undefined	1	1	X	429	25	Chile Transparente		
6	6	http://www.mineduc.cifra...	http://www.cdc.gob.cl/	1	Chile	undefined	1	1	X	1	9	GobiernoTransparente - Ge...		
7	7	http://transparencia.rede...	http://www.cdc.gob.cl/	1	Chile	gov	1	1	X	32	0			
8	8	http://www.portalde TRANSP...	http://www.cdc.gob.cl/	1	Chile	undefined	1	1	X	0	32	Portada - Portal de Trans...	Portal de transparencia d...	Portal de Transparencia d...
9	9	http://www.educatranspare...	http://www.cdc.gob.cl/	1	Chile	undefined	1	1	X	125	53	EducaTransparencia Insc...		
10	10	http://www.uchile.cl/trans...	http://www.cdc.gob.cl/	1	Chile	undefined	1	1	X	0	35	Transparencia - Universidad de Chile, uni...	Universidad de Chile, uni...	Universidad de Chile, ins...

Figura 46. Módulo de visualización de resultados. Extraído de VOSON.

Antes de generar las bases de datos de análisis es probable que sea necesario depurar los resultados obtenidos en la base de datos original con el fin, por ejemplo, de refinar el agrupamiento de páginas que VOSON realiza automáticamente porque comparten el dominio en su URL, eliminar páginas no relevantes para el análisis o separar algunas páginas agrupadas automáticamente en dominios como Wikipedia que contienen artículos de distinta temática.

Para hacer esto debemos abrir la base de datos “*Type: voson*” que contiene los resultados totales del rastreo y utilizar la opción “*Preferences/Node Pre-process/Database (Node)*”. En este módulo podemos agrupar páginas utilizando la sección “*Pagegrouping*” donde simplemente introduciremos los dominios que queremos agrupar separados por una coma; Podemos eliminar páginas de los resultados de análisis utilizando la opción “*Pruning*” donde introduciremos los dominios que queremos eliminar del análisis, separados por una coma. O podemos separar páginas agrupadas utilizando la opción “*Preserving*” e introduciendo las URLs de las páginas que no queremos que sean agrupadas por compartir el mismo dominio.

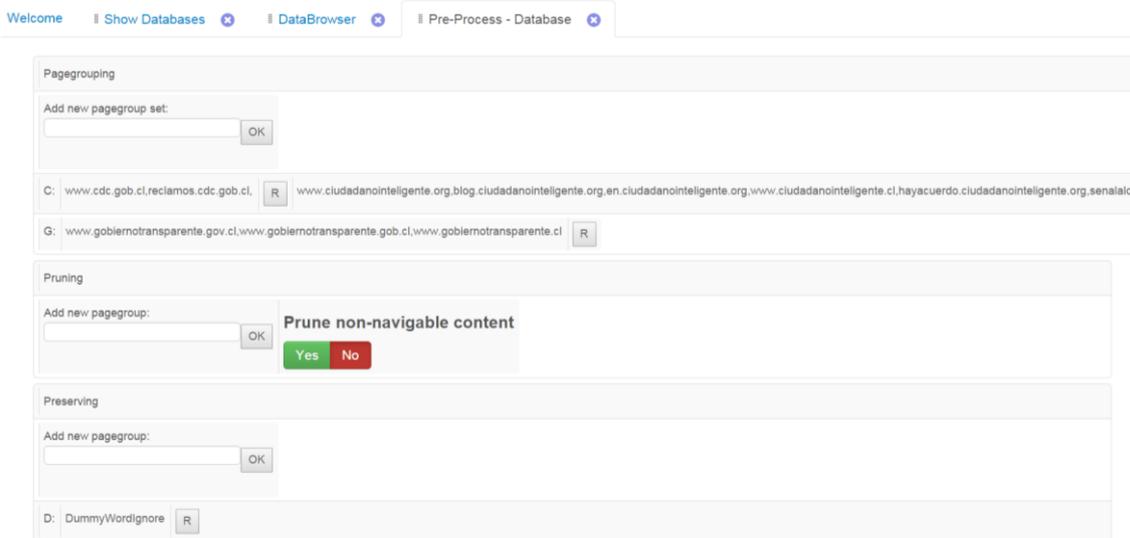


Figura 47. Módulo de preprocesamiento de datos. Extraído de VOSON.

VOSON nos permite además crear nuevas categorías en las que agrupar los resultados obtenidos en el análisis. A través de la opción “coding” podemos generar categorías en la base de datos principal de los resultados para posteriormente, en la base de datos de análisis, asignar los resultados a las nuevas categorías creadas. No obstante, el módulo de gestión de bases de datos presenta muchas limitaciones de usabilidad y este proceso puede resultar muy costoso en función del número de resultados que contenga nuestra base de datos.

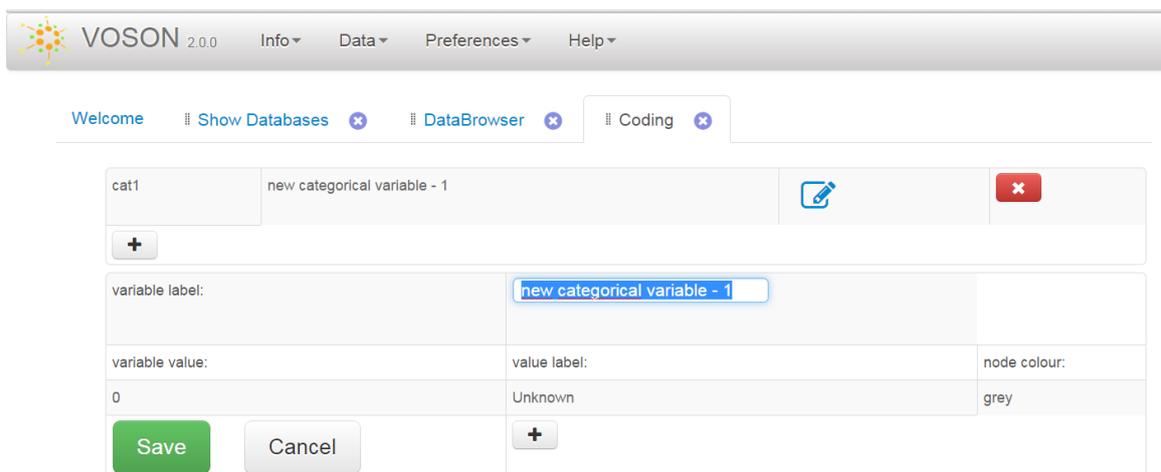


Figura 48. Módulo de codificación. Extraído de VOSON.

VOSON también ofrece la posibilidad de dividir la red absoluta en subredes utilizando cualquiera de los criterios de organización que hemos señalado previamente. A través de la opción *crosstabs – composition* podremos crear subredes, por ejemplo, de los clusters identificados al aplicar el algoritmo Newman – Girvan. Esta división en subredes nos permitirá extraer medidas de análisis de redes sociales de las subredes creadas o llevar a cabo análisis de los actores en relación a las subredes y no únicamente de la red absoluta.

Una vez hemos creado la bases de datos de análisis con los criterios necesarios para nuestro análisis y hemos depurado el agrupamiento de páginas y eliminado los resultados no relevantes podemos visualizar los resultados del análisis como mapa de redes utilizando la opción “*Analysis/Maps*”.

Este módulo nos permite elegir entre cuatro formas de visualización de mapas de redes: “*Minimun spanning tree*” que nos ofrece el camino más corto entre un nodo de la red y el resto de nodos con los que enlaza directa o indirectamente, “*Complete Network*” nos ofrece una representación completa de los nodos y enlaces que existen en la red, “*Concept*” que nos ofrece una representación de la relación entre los conceptos que contiene la red y finalmente la opción “*Hierarchy*” que nos ofrece una representación de los nodos organizados por su nivel de importancia dentro de la red.

En el formato *Minimum spanning tree* podemos configurar la representación de la red eligiendo mostrar enlaces entrantes o salientes, el nivel de separación entre sitios, el color de los nodos en función de varios criterios, la representación de los enlaces como dirigidos o no dirigidos, añadir al gráfico información sobre los nodos como el nombre del sitio o la URL, y finalmente el tamaño de los nodos en función del número de enlaces entrantes, salientes, etc. Vemos un ejemplo a continuación en la Figura 49.



Figura 49. Ejemplo de visualización de redes en formato *minimum spanning tree* generado con VOSON.

En el formato *Complete Network* podemos ver la red con todos los nodos y enlaces analizados. En este formato de representación podemos configurar el color de los nodos en relación a su número de enlaces entrantes, salientes, tipo de dominio, por clusters, etc. Los enlaces pueden ser representados como enlaces dirigidos, no dirigidos o no ser representados. Podemos incluir información sobre los enlaces como la URL o el nombre del sitio. Modificar el tamaño de los nodos en función del número de enlaces entrantes o salientes, de su autoridad para lo que se utiliza el algoritmo HITS²⁸, del nivel de intermediación (*betweenness*)²⁹, del nivel cercanía (*closeness*)³⁰, etc.

²⁸ El algoritmo HITS (*Hypertext Induced Topic Selection*) es un algoritmo diseñado por Jon Kleinberg para valorar la importancia de una página web. HITS usa dos indicadores para hacer esta valoración, definiendo recursivamente cada uno a partir del otro: el *authority*, que valora la calidad de una página como recurso de información usa para su cálculo una suma ponderada de valores *hub* de los enlaces que apuntan hacia esta página. El *hub*, determina la calidad de una información en función de los enlaces que tiene a otras páginas; este se calcula como una suma ponderada de valores *authority* de las páginas a las que apunta ésta https://es.wikipedia.org/wiki/Algoritmo_HITS

²⁹ *Betweenness centrality* cuantifica el número de veces que un nodo actúa como puente en el camino más corto entre otros dos nodos. <https://en.wikipedia.org/wiki/Centrality>

³⁰ *Closeness centrality* indica la centralidad de un nodo en una red y se calcula en función del número total de enlaces que separan un nodos de otros en la red. Se considera que un nodo

Al seleccionar este tipo de representación aparecerá un panel de control llamado “FDG Control” que nos permite seleccionar entre dos algoritmos de representación de redes: el algoritmo LinLog de Noack³¹ o el algoritmo Fruchterman y Reingold³². En este panel podemos además eliminar los nodos aislados de la red y elegir el número de iteraciones que lleva a cabo el algoritmo seleccionado con un máximo de 100 iteraciones. Junto al panel FDG Control encontraremos otro panel llamado FDG Legend donde veremos la leyenda de los colores que aparecen en el gráfico. Esta variará en función de los parámetros de configuración que hayamos seleccionado.

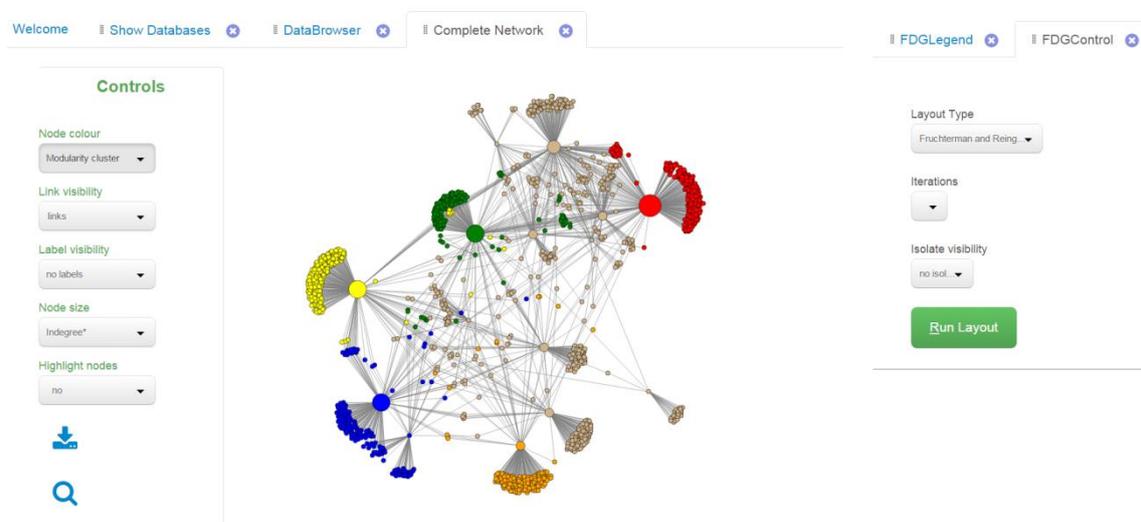


Figura 50. Ejemplo de visualización de redes en formato completo generado con VOSON.

con un alto grado de intermediación central (*Betweenness centrality*) tiene una gran influencia en la transferencia de información en la red. <https://en.wikipedia.org/wiki/Centrality>

³¹ LinLog es un algoritmo de energía diseñado por Andreas Noack que ofrece la posibilidad de minimizar la energía en grafos dirigidos por fuerza. El modelo Lin Log se basa en la idea de que la menor cantidad de energía se consigue minimizando el radio entre la distancia media de los enlaces y la distancia media entre los nodos.

<http://www.cs.columbia.edu/~itsik/software/ShareViz/LinLogModel.html>

³² Fruchterman-Reingold es un algoritmo para dibujar grafos dirigidos por fuerzas que fue desarrollado por Thomas Fruchterman y Edward Reingold en 1991. Se basa en el algoritmo previo de Eades y cuantifica la fuerzas atractivas entre vértices cercanos y las fuerza repulsivas entre todas las parejas de vértices. <http://cs.brown.edu/~rt/gdhandbook/chapters/force-directed.pdf>

En el formato *Concept* podemos ver una representación gráfica de la relación entre los conceptos extraídos de las páginas web analizadas. En este tipo de representación podemos seleccionar el tipo de conceptos que se representan pudiendo ser estas palabras clave, conceptos extraídos del cuerpo del texto o parejas de palabras extraídas del cuerpo del texto. Podemos además definir el límite máximo de coocurrencias de un concepto, el límite de coocurrencias de un concepto por documento y definir el color de los nodos en función del número de enlaces recibidos, salientes, etc.

El formato *Hierarchy* nos permite representar la red jerárquicamente en función de diferentes criterios de categorización: número de enlaces entrantes, salientes, *HITS hub*, *HITS authority*, cercanía (*closeness*), intermediación (*betweenness*), etc. En esta forma de representación, como en las anteriores, podemos elegir si queremos visualizar los nodos aislados, colorear los nodos en función de los criterios ya citados, visualizar los enlaces como dirigidos, no dirigidos o no visualizarlos y mostrar o no la información adicional de los nodos.

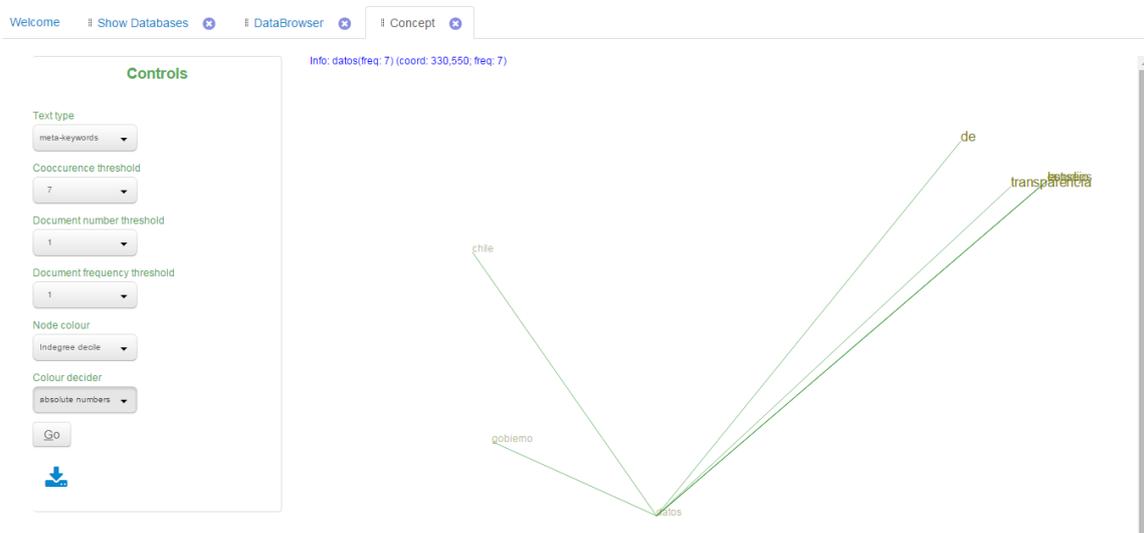


Figura 51. Ejemplo de visualización de redes de conceptos generado con VOSON.

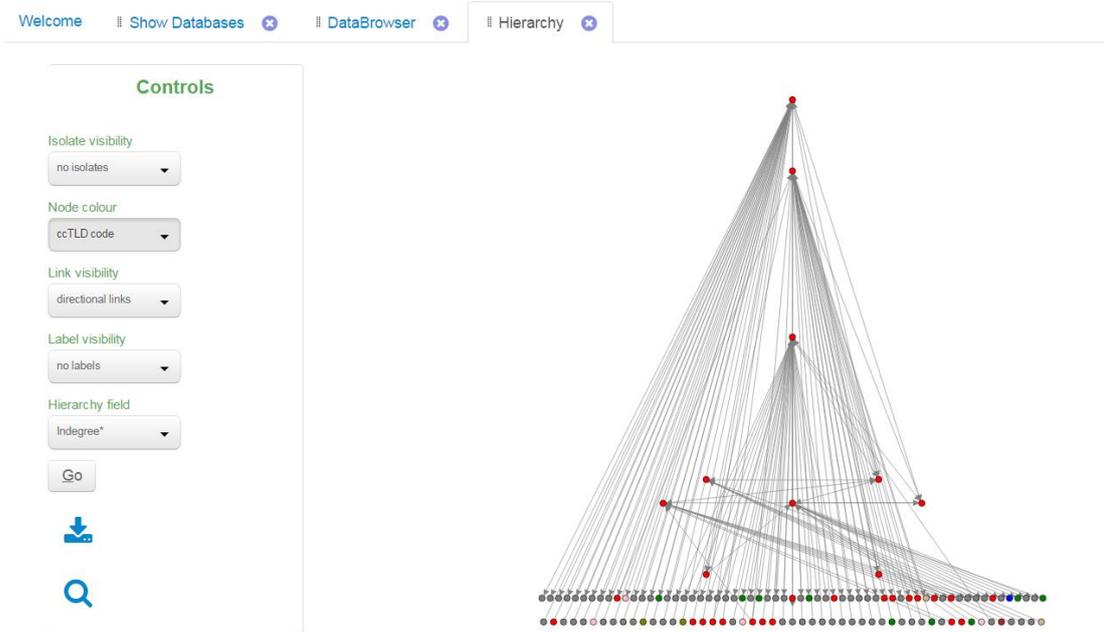


Figura 52. Ejemplo de visualización de redes en formato jerárquico generado con VOSON.

En los cuatro formatos de representación tenemos la opción de exportar la imagen en formato PNG y SVG. Además de exportar las imágenes podemos igualmente exportar las bases de datos en formato GraphML, Pajek y CSV.

3.3 Propuesta de modelo de calidad y evaluación de herramientas

En el apartado 3.2.1. hemos presentado las principales perspectivas y modelos de calidad relacionados con la evaluación de software. En el apartado 3.2.2 hemos descrito y caracterizado cuatro crawlers seleccionados por considerarse los más adecuados para el análisis de redes temáticas en la web. En este apartado 3.3 proponemos un modelo de calidad para la evaluación de software adaptado a las necesidades del análisis de redes temáticas en la web.

Para elaborar un modelo de calidad es necesario, en primer lugar, definir los objetivos relacionados con el uso de las herramientas que se van a evaluar. En nuestro caso vamos a evaluar web crawlers desde la perspectiva de un investigador que busca analizar redes temáticas en la web. Los objetivos de este análisis, como señalábamos

al inicio de este capítulo son: (1) analizar redes de hiperenlaces, (2) analizar los contenidos extraídos de esas redes (3) identificar los actores que participan en la red.

Además de cumplir con los objetivos señalados se espera que las herramientas seleccionadas tengan ciertos atributos de calidad deseables en cualquier software como por ejemplo una buena documentación que facilite el proceso de aprendizaje, un fácil uso de sus funcionalidades o su interoperabilidad con otras herramientas. Teniendo en cuenta todos estos aspectos, proponemos un modelo de calidad jerárquico con una serie de parámetros que serán definidos a través de diferentes indicadores y finalmente evaluados cualitativamente.

El modelo que proponemos utiliza la norma de calidad ISO 25010 como base para definir el conjunto de atributos e indicadores que consideramos relevantes para el análisis de calidad de una herramienta destinada al análisis de redes temáticas en la web. Recordemos que el modelo de calidad ISO 25010 propone un grupo de atributos destinados a la evaluación de calidad del producto software y un grupo destinado a la evaluación de la calidad en uso. Mientras que los atributos de calidad del producto software evalúan las características del producto directamente, los atributos de calidad en uso buscan evaluar las características del software en relación con diferentes grupos de interés del producto. En nuestro caso, y como ya hemos indicado, el modelo de calidad que proponemos se construye para satisfacer las necesidades de un investigador del ámbito de las Ciencias Sociales.

En el modelo que proponemos no haremos esa distinción entre atributos de calidad del software y atributos de calidad en uso, sino que haremos una selección de los atributos más relevantes para un usuario investigador y los unificaremos bajo el nombre de atributos de calidad de uso, considerando que estos atributos son generales para cualquier software. Los vemos a continuación la Figura 53.



Figura 53. Propuesta de modelo de evaluación de la calidad general en uso de crawlers. Elaboración propia.

Para evaluar las características específicas de las cuatro herramientas seleccionadas, que recordemos son crawlers destinados al rastreo parcial de la web, proponemos un segundo grupo de atributos relacionados con las funcionalidades propias de este tipo de software al que llamaremos atributos de calidad funcional. Las funcionalidades que se incluyen en el modelo de calidad funcional no pretenden hacer una evaluación exhaustiva de las herramientas, sino evaluar su adecuación a las necesidades del análisis de redes temáticas en la web.

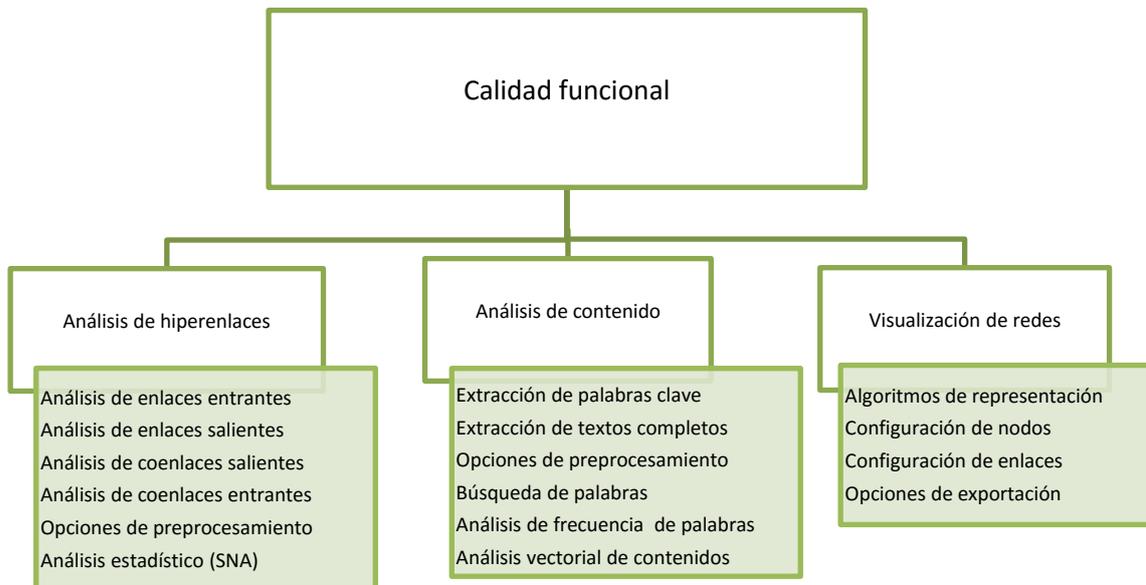


Figura 54. Propuesta de modelo de evaluación de la calidad funcional en crawlers. Elaboración propia

Para aplicar el modelo de calidad que acabamos de presentar a las herramientas seleccionadas crearemos dos tablas: la Tabla 10 servirá para evaluar los atributos de calidad general en uso; y la Tabla 11 servirá para evaluar los atributos funcionales de las herramientas. Estas dos tablas toman la información obtenida en el proceso de caracterización de las herramientas que hemos desarrollado previamente y utiliza el método de evaluación cualitativa indicando en ocasiones la existencia o no de las opciones señaladas por los indicadores del modelo de calidad y en los casos en que sea pertinente un método de evaluación cualitativa con diferentes grados de cumplimiento o adecuación en relación al indicador estudiado, siendo estos generalmente: buena, mala y regular o alto bajo y medio. Lo vemos a continuación en las Tablas 3 y 4.

Atributos	Indicadores	IssueCrawler	SocSciBot	Webometric Analyst	VOSON
Eficiencia	Comportamiento temporal	En función de la demanda	Variable en función del ancho de banda	Variable en función del ancho de banda	En función de la demanda y del tipo de suscripción
	Utilización de recursos	Procesamiento en la nube	Bajo consumo de recursos. Alto consumo de ancho de banda.	Bajo consumo de recursos. Alto consumo de ancho de banda.	Procesamiento en la nube
	Capacidad	Limitada	Ilimitada	Limitada gratuita – Ilimitada de pago	Limitada al tipo de suscripción – Ilimitada de pago
Compatibilidad	Coexistencia	Todos los navegadores	S.O. Windows	S.O. Windows	Firefox – Suscripción gratuita Todos los navegadores – Otras suscripciones
	Exportación	.svg, .jpg, .png, .tiff, .pdf, .html, .net, .xml, .txt, .gefx	.net, .txt	.net, .jpg, .txt	.png, .svg, .gml, .net, .csv
Usabilidad	Documentación	Regular	Regular	Regular	Buena
	Aprendizaje	Fácil	Medio	Difícil	Medio
	Operabilidad	Alta	Media	Media	Alta
	Estética	Buena	Mala	Regular	Muy buena
	Accesibilidad	Regular	Mala	Mala	Regular
Seguridad	Confidencialidad	Mala	Regular	Regular	Buena
	Responsabilidad	-	Alta	Alta	Media
Portabilidad	Instalación	-	Media	Fácil	-
	Coexistencia	-	Media	Media	-
	Reemplazo	-	Medio	Fácil	-

Tabla 10. Evaluación de la calidad general en uso de IssueCrawler, SocSciBot, Webometric Analyst y VOSON. Elaboración propia.

Atributos	Indicadores	IssueCrawler	SocSciBot	Webometric Analyst	VOSON
Análisis de hiperenlaces	Enlaces entrantes	x	x	✓	✓
	Enlaces salientes	✓	✓	✓	✓
	Coenlaces salientes	✓	x	x	✓
	Coenlaces entrantes	x	x	✓	✓
	Análisis estadístico (SNA)	✓	✓	✓	✓
	Opciones de preprocesamiento	x	Archivos de texto	Archivos de texto	Base de datos
Análisis de contenidos	Palabras clave	x	✓	x	✓
	Textos completos	x	✓	x	✓
	Opciones de preprocesamiento	x	✓	✓	✓
	Búsqueda de palabras	x	✓	x	✓
	Análisis de frecuencia de palabras	x	✓		✓
	Análisis vectorial de contenidos	x	x	x	✓
Visualización de redes	Algoritmos de representación	Indegree centrality	Fruchterman-Reingold	- Fruchterman-Reingold - Kamada Kawai	- Fruchterman-Reingold - LinLog de Noack
	Configuración de nodos	✓	✓	✓	✓
	Configuración de enlaces	✓	✓	✓	✓
	Opciones de exportación	.svg, .jpg, .png, .tiff, .pdf, .html, .net, .xml, .txt, .gefx	.net, .txt	.net, .jpg, .txt	.png, .svg, .gml, .net, .csv

Tabla 11. Evaluación de la calidad funcional de IssueCrawler, SocSciBot, Webometric Analyst y VOSON. Elaboración propia.

A la vista de los resultados de la evaluación podemos comprobar que las cuatro herramientas seleccionadas presentan una Calidad General de Uso media o buena siendo la más destacada en este aspecto de la evaluación VOSON del Australian Demographic and Social Research Institute de la Australian National University. IssueCrawler tiene la gran debilidad de que los datos extraídos y analizados con la herramienta se muestran públicamente en el archivo de la aplicación por lo que quedaría descartada en investigaciones que requieran de confidencialidad antes de la publicación de los resultados. SocSciBot es una herramienta que se puede considerar correcta en términos de calidad general pero tiene el gran inconveniente de no contar con una documentación completa lo que trae aparejado un largo proceso de aprendizaje basado en la prueba y error. Webometric Analyst por su parte, muestra la misma debilidad en la documentación de la herramienta, lo que provoca un largo proceso de aprendizaje que se acentúa aún más debido a la gran complejidad de la herramienta y a las numerosísimas utilidades que incorpora.

En cuanto a la evaluación de la calidad funcional podemos afirmar que VOSON se encuentra por encima del resto de herramientas y que cuenta con todas las funcionalidades necesarias para desarrollar el análisis de redes temáticas en todas sus fases. IssueCrawler no permite analizar enlaces entrantes que en el análisis de redes temáticas en la web, y en la mayoría de investigaciones de similares, resulta imprescindible. El mismo problema presenta SocSciBot, sin embargo, su uso combinado con Webometric Analyst puede aportar las mismas o incluso más utilidades de las que ofrece VOSON. La principal debilidad del uso combinado de SocSciBot y Webometric Analyst frente a VOSON es que esta última permite llevar a cabo análisis de enlaces entrantes y enlaces salientes en una única red lo que facilita notablemente el desarrollo del análisis y la interpretación de los resultados. Por todo ello, la herramienta que hemos considerado más adecuada para llevar a cabo cualquier investigación de análisis de redes temáticas en la web y esta investigación en particular, es VOSON.

4. DISEÑO Y APLICACIÓN DE UN SISTEMA DE ANÁLISIS DE REDES TEMÁTICAS EN LA WEB

El tercer objetivo de nuestra investigación se aborda en la primera parte de este Capítulo 4, y consiste en **Diseñar un Sistema de Análisis de Redes Temáticas en la Web**.

El cuarto objetivo se aborda en la segunda parte de este Capítulo 4, y consiste en **Aplicar el Sistema de Análisis de Redes Temáticas en la Web al caso de estudio de la Transparencia en Chile**.

4.1. Diseño de un Sistema de Análisis de Redes Temáticas en la Web

Como ya se ha señalado previamente, consideramos que el análisis de redes temáticas, para ser completo, debe desarrollarse en tres fases que corresponden al análisis de los tres componentes que forman parte de una red temática y que hemos señalado y definido en el Capítulo 2, estos son: (1) las redes de hiperenlaces, (2) los temas y (3) los actores.

El **Sistema de Análisis de Redes Temáticas en la Web** que proponemos se divide igualmente en tres fases que corresponden con los tres componentes señalados:

- **Análisis de redes de hiperenlaces**, entendiendo una red de hiperenlaces como un sistema dinámico que hace posible el flujo de información entre diferentes sitios web y a través del cual se establecen relaciones de comunidad entre los actores que los crean.
- **Análisis de temas**, entendiendo temas como conjuntos de ideas, teorías o conceptos que generan controversia entre grupos de actores políticos, sociales o económicos.
- **Análisis de actores**, entendiendo actores como individuos u organizaciones del ámbito público o privado que tengan quejas, recursos u oportunidad de influir en un tema de discusión.

4.1.1. Fases y componentes del análisis de redes de hiperenlaces

La primera fase del Sistema de Análisis de Redes Temáticas es la que corresponde al análisis de las redes de hiperenlaces que subyacen a las redes temáticas en la web. Para analizar esas redes de hiperenlaces, en primer lugar, tenemos que ser capaces de identificarlas y para conseguirlo buscaremos los enlaces que unas páginas dirigen a otras utilizando para ello crawlers.

Recordemos que los crawlers son herramientas capaces de buscar, recuperar y analizar los documentos de la Web y que sus características varían en función de los objetivos que persiguen. Existen crawlers no preferenciales que recuperan páginas de la web sin tener en cuenta ningún criterio de selección y crawlers preferenciales que seleccionan las páginas web que recuperan en base a criterios como la actualización, el tamaño o el número de enlaces que reciben.

El análisis de redes temáticas requiere de un rastreo preferencial de la web acotado a un área temática, pues recordemos que consideramos una red temática a un conjunto de actores que a través de una red de hiperenlaces participan en un tema de discusión. La medida en que las redes temáticas representan una discusión tiene que ver con el reconocimiento mutuo de los actores y de las diferentes posiciones que defienden. La participación de actores de diferentes ámbitos es fundamental para la existencia de una red temática pues de lo contrario estaríamos ante un discurso parcial y no ante un tema de discusión.

El método más habitual para desarrollar un rastreo preferencial definido por un tema, y el que utilizan los cuatro crawlers que hemos estudiado en el Capítulo 3, comienza por seleccionar un grupo de sitios web relevantes en el área temática que queremos analizar. Estos sitios web actuarán como semillas y servirán de punto de origen a la herramienta para lanzar el rastreo.

a) Selección de semillas

La selección de esos sitios que actuarán como semillas debe ser representativa de todos los ámbitos que intervienen en el tema de discusión sin privilegiar, por ejemplo, el ámbito

gubernamental .gob, el ámbito no gubernamental, .org o el ámbito académico, .edu. Para conseguir un grupo de semillas que represente a todos los ámbitos de la red temática existen diferentes estrategias que presentamos a continuación.

- **Consultas en buscadores:** consiste en realizar consultas de palabras o frases relacionadas con el tema en cuestión en los principales buscadores y seleccionar los 10 o 20 resultados más destacados. Esta estrategia es simple y poco costosa en términos de tiempo pero tiene el inconveniente de que los buscadores ofrecen resultados de la totalidad de la web y nosotros estamos interesados en una parte de ella. Los buscadores además utilizan criterios de ordenación en sus índices que no corresponden con nuestros criterios de análisis y pueden dejar fuera de sus resultados páginas que puedan ser de interés en nuestro análisis.
- **Directorios:** consiste en buscar directorios, listas de distribución o cualquier tipo de índice de organizaciones relacionadas con el tema que queremos investigar. El problema de esta opción es que en función del tema puede que no existan esos índices o que los índices existentes sean parciales o solo representativos de una parte de la red.
- **Medios de comunicación:** consiste en buscar informaciones relacionadas con el tema en los principales medios de comunicación y recopilar los nombres de los actores que aparecen. El inconveniente de esta estrategia es que no podemos estar seguros de que todas las partes implicadas en un tema de discusión aparezcan representadas en los medios de comunicación.
- **Consulta de expertos:** consiste en preguntar a uno o varios expertos relacionados con el área temática en cuestión. Tiene el inconveniente del coste temporal que supone y el riesgo de no poder evitar la subjetividad de los encuestados.
- **Rastreo de co-enlaces entrantes:** se desarrolla a partir de un grupo inicial de semillas que debemos obtener utilizando las técnicas descritas previamente y consiste en identificar aquellos sitios web que envían enlaces hacia dos de las páginas seleccionadas en el grupo de semillas. El análisis de co-enlaces entrantes tiene la propiedad de ofrecer como resultado sitios similares a los introducidos como

semillas por lo que puede ser utilizado para ampliar un grupo de semillas inicial y mitigar así las limitaciones que hemos señalado en las técnicas previas.

Para conseguir un grupo de semillas relevantes y que sea representativo de los diferentes ámbitos que participan en un tema de discusión proponemos combinar varias de estas estrategias. La consulta en buscadores combinada con la consulta de directorios o medios de comunicación son las combinaciones que requieren de menos recursos al investigador. En caso de que no existan directorios relacionados con el tema de discusión será recomendable consultar a expertos o revisar los medios en busca de actores relevantes. En el caso de que el grupo de semillas obtenido por estos medios sea poco numeroso o consideremos que es parcial o incompleto proponemos ampliar el grupo de semillas aplicando un análisis de coenlaces. Este proceso de análisis de coenlaces puede presentar resultados poco relevantes y debe hacerse siempre acompañado de un análisis cualitativo.

b) Análisis de enlaces

Una vez seleccionado el grupo de páginas que actuarán como semillas debemos buscar las páginas que presentan algún tipo de enlace con estas o los enlaces que existen entre ellas. Como ya hemos señalado en capítulos anteriores, existen diferentes tipos de enlaces cuyo análisis ofrece diferentes tipos de información.

Recordemos que un hiperenlace es la porción del código fuente de una página web (elemento <a> más atributo href) que activa un nuevo destino de navegación si el usuario hace clic sobre el mismo. Pero recordemos también que una página web A puede contener la URL de una página web B sin contener necesariamente un hiperenlace dirigido a ella y que esto es lo que consideramos una cita de URL o "*URL citation*".

Como explicábamos en capítulos anteriores, los grandes buscadores han retirado la mayoría de sus servicios de información sobre hiperenlaces lo que ha llevado a algunos investigadores a utilizar este tipo de citas de URL como alternativa a los hiperenlaces. Teniendo en cuenta los resultados obtenidos por Thelwall en sus investigaciones sobre la validez del uso de las citas de URL como alternativa a los hiperenlaces (Thelwall, 2012), proponemos utilizar tanto hiperenlaces como citas de URL.

Thelwall planteaba además la utilización de menciones, es decir, la aparición del nombre de un actor, individuo u organización, en una página web, pero como él mismo señala estas pueden representar relaciones pero no tienen por qué tener las propiedades características que se ha demostrado tienen los hiperenlaces. Por todo ello no incorporamos esta técnica en nuestra metodología y recomendamos la utilización de hiperenlaces siempre que sea posible y la utilización de citas de URL cuando no sea posible utilizar hiperenlaces.

Además de las diferencias entre hiperenlaces y citas de URL existen diferentes tipos de enlace en función de sus características. Existen diferentes clasificaciones pero una adaptada a las necesidades del análisis de redes temáticas sería la que incluye enlaces salientes (*outlinks*), enlaces entrantes (*inlinks*), enlaces recíprocos (*recíprocal links*), coenlaces salientes (*co-outlinks*) y coenlaces entrantes (*co-inlinks*). A continuación los definimos brevemente y explicamos los usos o valores que se le da a cada uno de ellos. Todos los gráficos que siguen a continuación son de elaboración propia.

Los **enlaces salientes** son los enlaces que un sitio web dirige a otras páginas.

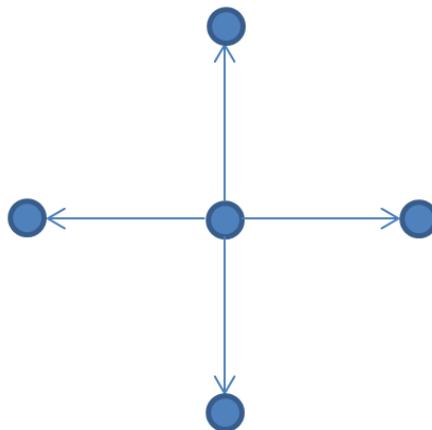


Figura 55. Ejemplo de nodo con enlaces salientes. Elaboración propia.

Se suele considerar que esas páginas que reciben los enlaces son sitios relevantes para la organización porque ofrecen una información valiosa o bien porque existe algún tipo de

alianza o relación. Recordamos que los enlaces pueden servir para proveer información, construir redes de refuerzo, intercambiar audiencia o amplificar mensajes.

Los **enlaces entrantes** son los enlaces que un sitio web recibe desde otras páginas.

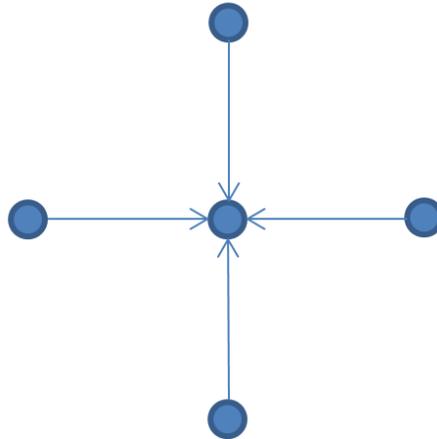


Figura 56. Ejemplo de nodo con enlaces entrantes. Elaboración propia.

Estos enlaces muestran los sitios que consideran relevante el sitio que recibe el enlace, bien porque existe una relación con el tema que trata bien porque consideran que el sitio que recibe el enlace contiene información relevante. Los enlaces entrantes que recibe un sitio web se consideran una medida de la importancia o capacidad de influencia de ese sitio y han sido uno de los criterios fundamentales de los grandes buscadores en la creación de sus rankings.

Los **enlaces recíprocos** son los enlaces mutuos que se hacen entre dos sitios web.



Figura 57. Ejemplo de nodos con enlaces recíprocos. Elaboración propia.

Este tipo de enlaces suele asociarse con un alto grado de relación o alianza entre las organizaciones. Cuanto mayor sea el número de enlaces recíprocos mayor probabilidad hay de que exista una forma de colaboración, alianza o pertenencia entre las dos organizaciones.

Los **coenlaces salientes** son un tipo de enlaces indirectos entre dos sitios que envían un enlace a un tercer sitio.

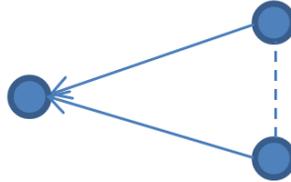


Figura 58. Ejemplo de nodos conectados indirectamente por coenlace saliente. Elaboración propia.

De igual forma que sucede con las parejas de citas (*coupling*) en el ámbito de las Ciencias de la Información, se considera que dos sitios que realizan un enlace hacia un tercero presentan cierto grado de similitud aunque no exista un enlace directo entre ellos. Cuanto mayor es el número de coenlaces mayor será el grado de similitud.

Los **coenlaces entrantes** son una forma de enlace indirecto que existe entre dos sitios que reciben un enlace desde un tercer sitio. Se considera que el sitio que realiza los enlaces también comparte cierto grado de similitud con los dos sitios enlazados y que ese grado de similitud será mayor cuanto mayor sea el número de enlaces que envía hacia esos sitios.

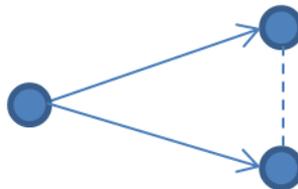


Figura 59. Ejemplo de nodos conectados indirectamente por coenlace entrante. Elaboración propia.

De la misma manera que en Ciencias de la Información ocurre con las cocitas (*co-citation*) se considera que dos sitios que son enlazados por un tercero presentan cierto grado de

similitud aunque no exista un enlace directo entre ellos. Este tipo de enlace es el que más se ha utilizado en la identificación y análisis de comunidades y redes temáticas en la web (Thelwall, 2012), no obstante, nosotros proponemos identificar además los sitios que presentan enlaces recíprocos con el grupo de semillas.

El análisis de coenlaces nos permite identificar actores con un mayor grado de similitud respecto a las semillas y el análisis de los enlaces recíprocos nos permite identificar a los sitios que tienen una relación fuerte ya sea de colaboración, de alianza o de pertenencia con las semillas. Consideramos que ambas informaciones son relevantes para el análisis de redes temáticas por lo que proponemos utilizar ambos tipos de enlaces como filtro en la depuración de la red inicial.

VOSON ofrece la posibilidad de crear subredes teniendo en cuenta únicamente los sitios que presentan coenlaces o enlaces recíprocos con las semillas. Al seleccionar la opción “*create seeds y seeds + important sites*” obtendremos una subred formada únicamente por las semillas introducidas en la herramienta y una segunda red formada por los sitios que presentan dos o más enlaces con las semillas, los sitios que tienen un enlace recíproco con las semillas y las propias semillas. VOSON considera tanto los coenlaces entrantes como los salientes, dos tipos de enlace que como decíamos antes han sido ampliamente empleados para detectar sitios similares (Rogers, 2002; Thelwall, 2006; Ackland 2005).

En la subred de coenlaces y enlaces recíprocos aplicaremos un agrupamiento (clústering) para detectar las comunidades que existen dentro de la red. El clústering forma parte de las técnicas de análisis de redes sociales que explicamos a continuación.

c) Análisis de redes sociales

Como señalábamos en el Capítulo 1, cuando entendemos los hiperenlaces como una forma de relación, podemos utilizar los métodos de análisis de redes sociales para describir los patrones de conexiones o desconexiones que se observan en una red de sitios web (Park, 2003: 48; Jackson, 1997: 292). El análisis de redes sociales es un ámbito de conocimiento amplio y no pretendemos aquí presentar un marco teórico

completo, sino solo explicar algunas de las medidas que emplearemos en el análisis de redes temáticas en la web.

El análisis de redes sociales puede desarrollarse desde dos perspectivas, la que analiza la red en su conjunto o la que analiza individualmente los nodos que forman parte de ella. En esta parte de la metodología dedicada al análisis de redes de hiperenlaces propondremos una serie de medidas relacionadas con el análisis de la red en su conjunto. No obstante, antes de aplicar estas medidas debemos asegurarnos de que todos los actores de la red han sido correctamente considerados y que no existen duplicaciones o divisiones de páginas que pertenecen a un mismo actor. En el apartado 4.1.2. Metodología para el análisis de actores, explicamos los procesos previos que se deben llevar a cabo para identificar claramente a los actores que forman parte de la red. Las medidas de análisis de redes sociales que se relacionan con el estudio de los nodos de una red también se presentarán ese apartado.

Para analizar las redes de hiperenlaces en su totalidad las medidas que utilizaremos serán:

- La modularidad de clústers (*clúster modularity*),
- La densidad (*density*),
- La centralidad (*degree centrality*),
- La intermediación (*betweenness centrality*)
- La cercanía (*closeness centrality*).

El **agrupamiento en clústers** o **clústering** hace referencia a la identificación de comunidades dentro de una red. Para calcular o buscar comunidades dentro de una red existen diferentes métodos y algoritmos, pero por su relevancia, nos presentaremos el algoritmo Newman-Girvan que es el empleado por la herramienta VOSON que utilizaremos en nuestro análisis.

Este algoritmo utiliza el concepto de intermediación (*betweenness*) aplicado a los enlaces de la red (*edges betweenness*), un indicador que determina el número de rutas o caminos posibles entre dos nodos que utilizan como puente un tercero. El algoritmo Girvan-Newman detecta comunidades eliminando progresivamente los enlaces de la red. A

medida que elimina los enlaces con un nivel de intermediación mayor van haciéndose visibles los clústers o comunidades que existen en la red.

Para aplicar este algoritmo, en primer lugar debemos calcular el nivel de intermediación de todos los enlaces de la red para posteriormente comenzar a eliminar aquellos enlaces que presentan mayor nivel de intermediación. Una vez eliminados esos enlaces volveremos a recalcular el nivel de intermediación de todos los enlaces y repetiremos este ciclo hasta que todos los enlaces sean eliminados. En ese proceso irán apareciendo grupos de nodos desconectados de otros que se podrán considerar como clústers o comunidades.

La **densidad** de una red nos muestra el nivel de conectividad que existe entre los nodos que forman parte de ella y para calcularlo debemos dividir el número de relaciones que existen en la red entre el número de las relaciones que podrían existir. Esta medida tendrá una exactitud relativa en nuestro caso ya que la densidad de las relaciones que existan en la red variará en función del tipo y número de semillas que consideremos inicialmente. Es de esperar que la densidad de la red aumente si aumentamos el valor de conectividad mínimo que utilizamos como límite en la selección de semillas así que el resultado final de conectividad de la red dependerá de una decisión propia y en cierto grado arbitraria. Sí puede ser de gran valor en la comparación de redes siempre y cuando utilicemos el mismo criterio en ambas para establecer los límites de relevancia mínima en la selección de semillas.

La **centralidad** o **centralización**³³ (*centrality* o *centralization*) de una red nos muestra el nivel relativo de centralidad de la red en su conjunto en comparación con el nodo que presenta mayor grado de centralidad. Para calcularla debemos sumar las diferencias en el grado de centralidad que existen entre el nodo con mayor nivel de centralidad y todos los demás nodos de la red y dividirlo por el mayor valor posible de diferencia entre los nodos. Esta medida nos muestra la tendencia que existe en la red a ser dominada por un nodo. Algunos autores (Leavit, 1951, Freeman, 1977) han defendido que la percepción de liderazgo y de eficiencia en la resolución de problemas que tienen los actores de una red aumenta cuando crece el grado de centralidad.

³³ En ocasiones, se utiliza el término centralización para hablar del nivel de centralidad de una red en su conjunto y diferenciarlo de centralidad que normalmente se refiere al nivel de centralidad de los nodos por separado.

La **intermediación** (*betweenness centrality*) en una red hace referencia al número de veces que un nodo actúa como puente entre dos nodos que no están conectados directamente. El grado de intermediación de una red en su conjunto se calcula, igual que sucedía con el nivel de centralidad, sumando las diferencias entre el nodo que presenta un mayor nivel de intermediación y todos los demás nodos, y dividiendo el valor obtenido entre el mayor número posible de intermediación posible en la red. Este valor puede interpretarse como el nivel de interconexión indirecta entre nodos a través de otros nodos de la red y se suele relacionar el nivel de control que tienen los nodos en los flujos de información que se producen en la red (Freeman, 1977).

La **cercanía** (*closeness centrality*) en una red hace referencia a la distancia que existe entre los nodos de la red. Para calcularla debemos sumar las diferencias que existen entre el nodo con mayor nivel de cercanía y dividirlo por el nivel máximo de cercanía que puede existir en la red. El nivel máximo de cercanía ocurre cuando todos los nodos están a un solo paso de distancia del resto de nodos. Esta medida suele ser interpretada como el nivel de cohesión que existe entre los nodos de la red y suele relacionarse como el nivel de afinidad o colaboración que existe entre los diferentes actores que participan en la red.

Existen otras medidas relacionadas con el análisis de redes sociales que se aplican a redes en su conjunto. La herramienta VOSON, por ejemplo, ofrece además medidas de centralidad basadas en el número de enlaces entrantes o del número de enlaces salientes. En la metodología de análisis de redes de hiperenlaces que proponemos utilizaremos únicamente las medidas estudiadas pues consideramos que ofrecen la información más relevante en relación con el análisis de redes temáticas en la web.

d) Visualización de redes de hiperenlaces

Para visualizar las redes de hiperenlaces que subyacen a las redes temáticas en la web debemos ser capaces de representar de manera gráfica los nodos que representan a los actores de la red y los enlaces que existen entre los diferentes actores. Como hemos visto previamente, los enlaces pueden ser de diferentes tipos: entrantes, salientes o recíprocos y deben ser claramente representados para diferenciar unos de otros. En el caso de los

actores lo veremos en el siguiente apartado: 4.1.2. Metodología para el análisis de actores, pero también pueden ser analizados en función de variadas características, como el número de enlaces que reciben, el nivel de autoridad o su grado de intermediación en la red. Todas estas características de los actores deben poder ser representadas en la visualización de redes.

Para conseguirlo utilizaremos la opción de visualización *Complete Network* que ofrece VOSON que nos permite crear un gráfico con todos los nodos y enlaces que forman parte de la red. En esta opción de visualización podemos modificar el tamaño de los nodos en relación a su número de enlaces entrantes, salientes o al nivel de autoridad que presentan. Podemos modificar el color de los mismos en función de su ámbito geográfico y de su ámbito de actividad para lo que VOSON utiliza el tipo de dominio.

El módulo de representación *Complete Network* de VOSON cuenta además con un panel "*FDG Control*" que nos permite seleccionar entre dos algoritmos de representación de redes: el algoritmo LinLog de Noack o el algoritmo Fruchterman-Reingold.

El algoritmo LinLog es un algoritmo de energía diseñado por Andreas Noack que ofrece la posibilidad de minimizar la energía en grafos dirigidos por fuerza. El modelo Lin Log se basa en la idea de que la menor cantidad de energía se consigue minimizando el radio entre la distancia media de los enlaces y la distancia media entre los nodos.

El algoritmo Fruchterman-Reingold es un algoritmo para dibujar grafos dirigidos por fuerzas que fue desarrollado por Thomas Fruchterman y Edward Reingold en 1991. Se basa en el algoritmo previo de Eades y cuantifica la fuerzas atractivas entre vértices cercanos y las fuerza repulsivas entre todas las parejas de vértices.

En nuestra metodología proponemos utilizar el algoritmo Fruchterman-Reingold que permite una visualización más clara ya que representa los nodos separados y permite visualizar claramente los enlaces que existen entre los diferentes nodos y el tipo de enlaces de cada uno de ellos. El algoritmo LinLog reúne los nodos en clústers muy agrupados y hace difícil la visualización de los enlaces entre nodos.

Para analizar la red proponemos utilizar varias visualizaciones: de la red absoluta, de la red de coenlaces y enlaces recíprocos y de los clústers que aparecen en la red. Todas ellas asociadas con las medidas de análisis de redes sociales previamente señaladas.

- La visualización de la red absoluta nos ofrecerá una representación del tamaño de la red, de la densidad de conexiones y de los principales actores que participan en ella con el número total de enlaces de cada uno. Este tipo de representación puede resultar indescifrable por el alto número de nodos y por la baja relación temática de algunos de ellos.
- La visualización de la red de coenlaces y enlaces recíprocos nos ofrece una representación refinada de los sitios que se pueden considerar más afines al grupo de semillas. En ella será posible analizar los actores que no pertenecen al grupo de semillas pero que se pueden considerar afines temáticamente por estar enlazados con dos o más de ellas o por presentar un enlace recíproco con alguna de las semillas.
- La visualización de los clústers de la red de coenlaces nos ofrece una representación de las comunidades existentes en la red. Este es un primer paso para analizar posteriormente las diferentes posiciones o marcos de referencia que existen en la red respecto al tema de discusión. La visualización individual de los clústers existentes nos permitirá comparar los diferentes niveles de densidad, centralidad, intermediación y cercanía que existe en cada comunidad lo que se podrá asociar con diferentes características de comportamiento de cada comunidad.

Fases y Componentes del Sistema de Análisis de Redes de Hiperenlaces

Selección de semillas

- Consultas en buscadores
- Directorios
- Medios de comunicación
- Consulta de expertos
- Rastreo de co-enlaces entrantes

Análisis de enlaces

- Enlaces entrantes
- Enlaces salientes
- Enlaces recíprocos
- Coenlaces entrantes
- Coenlaces salientes

Análisis de redes sociales

- Clústering
- Densidad
- Centralidad
- Intermediación
- Cercanía

Visualización de redes

- Red absoluta
- Red de coenlaces
- Clústers

4.1.2. Fases y componentes del análisis de actores

Una vez analizadas las redes de hiperenlaces que subyacen a las redes temáticas en la web, la segunda fase del análisis consistirá en analizar los actores que participan en esas redes. Para conseguirlo, en primer lugar tendremos que identificar claramente a los actores y asegurarnos de que cada actor es representado mediante un nodo en la red.

a) Identificación de actores

En ocasiones, un único actor puede estar representado por varios nodos en la red si por ejemplo cuenta con varios sitios web diferentes. Otras veces, el análisis de hiperenlaces no será capaz de agrupar en un único nodo dos dominios pertenecientes a un único sitio web. Pongamos un ejemplo: la organización “Ciudadano Inteligente” cuenta con un sitio web cuya URL es ciudadanointeligente.org pero también es propietario de la URL ciudadanointeligente.cl que redirige al sitio principal. En este caso, el crawler no sabrá que estas dos URLs pertenecen a un mismo actor y por tanto las analizará por separado y las

representará con dos nodos en la visualización de la red. Otra posibilidad es que el sitio web de una organización cuente con varios subdominios. La misma organización “Ciudadano Inteligente” cuenta con su sitio principal ciudadanointeligente.org pero también cuenta con el subdominio blog.ciudadanointeligente.org. En este caso, si no hemos configurado el crawler para agrupar todos los dominios que contienen la URL del sitio web principal, tampoco identificará que el subdominio en blog.ciudadanointeligente.org pertenece al sitio principal ciudadanointeligente.org.

La configuración del crawler es pues fundamental pero no siempre la mejor opción será agrupar todos los dominios que contienen la URL de un sitio web. Por ejemplo, en el caso de que queramos considerar los blogs dentro de nuestro análisis, nos encontraremos con que existen muchos blogs que utilizan las URLs que ofrecen sitios como Wordpress o Blogspot.

Es decir, podemos encontrarnos con URLs del tipo blogdetrasnparencia.blogspot.com o blogdetrasnparencia.wordpress.com. Si configuramos el crawler para agrupar todos los resultados que comparte la URL de un sitio perderemos los resultados de estos blogs que se agruparán en los resultados de wordpress.com o blogspot.com. Si utilizamos la herramienta VOSON será relativamente fácil agrupar resultados a través de la opción “*page grouping*” que nos permite modificar los parámetros que VOSON utiliza para crear grupos de páginas automáticamente.

De la misma forma que podemos necesitar agrupar páginas y subdominios en un único nodo que represente a un actor, podemos necesitar separar resultados de páginas que hayan sido agrupados pero en realidad representan a varios actores.

Normalmente los sitios web agrupan sus páginas en directorios que representan las diferentes secciones del sitio. Para conseguir agrupar todas las páginas de diferentes directorios en un único nodo que representa al actor que ha creado ese sitio necesitaremos configurar el crawler para agrupe automáticamente todas las páginas que contienen la URL del sitio en cuestión. Sin embargo, en algunos casos podemos encontrarnos con sitios que albergan páginas que representan a diferentes actores.

Pongamos un ejemplo. En el caso de Wikipedia, podemos encontrar una entrada relacionada con la organización *Ciudadano Inteligente* en la URL wikipedia.es/ciudadanointeligente. El crawler considera que todos los directorios del sitio Wikipedia representan a Wikipedia y por tanto los agrupa. En este caso no deberemos agrupar resultados, sino separarlos. Si utilizamos VOSON lo podremos hacer fácilmente utilizando la opción “*preserving*” que nos permite crear nuevos grupos de páginas a partir de cualquiera de las páginas obtenidas como resultado del rastreo.

Una vez hemos agrupado y separado las páginas de los resultados para representar correctamente a los diferentes actores que forman parte de la red debemos identificar su nombre y el ámbito al que pertenecen. Para identificar el nombre de la organización o individuo que hay detrás de un sitio web utilizaremos la información que el crawler extrae de las páginas en el rastreo. Si utilizamos VOSON, la herramienta utilizará la URL del sitio principal para dar nombre a los grupos de páginas que se han creado. Este puede ser modificado por el nombre de la entidad si lo consideramos conveniente pero normalmente la URL nos permitirá identificar el nombre del actor que hay detrás de ese sitio web. En los casos en que esto no sea posible podemos tratar de conseguirlo a través de la información que aparece en el atributo Palabras Clave (*Page Keywords*).

Para identificar el ámbito al que pertenece cada actor podemos utilizar el atributo *genericTLD* en el que encontraremos el tipo de dominio que utiliza cada página. Así, si el dominio es del tipo .org podremos inferir que ese actor participa en el tema de discusión desde el ámbito de las organizaciones no gubernamentales. Si el dominio es del tipo .gov el actor pertenecerá al ámbito gubernamental o si el dominio es del tipo .edu podremos decir que pertenece al ámbito académico. Este tipo de análisis presenta la dificultad de que no todas las organizaciones utilizan los tipos de dominios que se suelen asociar con estos tres ámbitos y existen organizaciones gubernamentales, por ejemplo, que pueden utilizar un tipo de dominio nacional como .cl o .es.

Una alternativa sería recurrir nuevamente a las palabras claves y a la descripción de la página que el crawler extrae de los atributos html de una página pero tampoco todos los sitios aportan siempre esta información. Para evitar todos los posibles errores sería necesario llevar a cabo un análisis cualitativo de las páginas pero esto sería muy costoso en el análisis de redes de gran tamaño.

Para identificar el ámbito geográfico de los actores podemos utilizar el atributo ccTLD de las páginas recuperadas pero este solo será útil en el caso de las páginas que utilizan un tipo de dominio geográfico como .cl o .es. Las páginas que utilicen un dominio de tipo .org, .com o .gov requerirán otro tipo de análisis. Nuevamente los atributos de palabras clave y descripción de la página nos permitirán en algunos casos resolver el problema pero como decíamos no todas las páginas aportan este tipo de información y si queremos evitar los errores deberemos analizar cualitativamente las páginas.

b) Análisis de redes sociales

Como decíamos en el apartado anterior 4.1.1. Metodología para el análisis de redes de hiperenlaces, el análisis de redes sociales se puede desarrollar desde la perspectiva de análisis de la red en su conjunto o analizando por separado los nodos que forman parte de la red.

Las medidas de análisis de redes sociales que utilizaremos para analizar los actores de una red temática serán el grado de centralidad (*degree centrality*) que se puede calcular en función de los enlaces entrantes (*indegree centrality*) y de los enlaces salientes (*outdegree centrality*), el grado de intermediación (*betweenness centrality*), el grado de cercanía (*closeness centrality*) y el valor de HITS en sus dos versiones *HITS authority* y *HITS hub*.

Estas medidas serán calculadas sobre la red absoluta, de la misma forma que la representación gráfica de sus relaciones, que plantearemos más adelante, considerará todos los enlaces y no únicamente los coenlaces. Esto se debe a que consideramos que en el estudio de los actores puede ser relevante obtener información sobre las relaciones de un actor no solo con otros actores afines temáticamente, sino también con actores menos similares o menos relacionados con el tema de discusión.

La **centralidad** (*degree centrality*) de un nodo se relaciona con el número de conexiones directas que presenta y se suele interpretar como una medida de importancia. Se considera que los actores más conectados en una red son los más importantes y los menos conectados son los menos importantes. Esto puede parecer evidente pero cuando estudiemos las medidas relacionadas con el prestigio de los actores, veremos que la

medida de la centralidad no es la única forma de analizar el prestigio de un actor y en función del tipo de red puede no ser la más adecuada.

Para calcular esta medida en gráficos no dirigidos, es decir aquellos que no diferencian entre enlaces entrantes y salientes, debemos sumar los enlaces del nodo en cuestión y normalizarla dividiéndolo por el número máximo de enlaces que podría presentar. En el caso de gráficos dirigidos, simplemente haremos el mismo cálculo teniendo en cuenta únicamente los enlaces entrantes o los enlaces salientes.

En nuestro caso es complicado calcular el valor normalizado de centralidad de un nodo porque el rastreo que llevan a cabo las herramientas que hemos estudiado contabiliza el número de enlaces que recibe un sitio desde otras páginas. Como ya hemos explicado, un único sitio puede albergar varias páginas y por tanto el cálculo de los enlaces que recibe un nodo puede considerar varios enlaces provenientes de un único nodo lo que hace imposible el cálculo normalizado de la centralidad.

La solución que proponemos a esta dificultad es considerar únicamente el número de enlaces entrantes o salientes que presenta un nodo y utilizar este número como medida de centralidad. Así, consideraremos que el nodo que recibe mayor número de enlaces entrantes es el nodo que presenta un mayor grado de centralidad entrante (*indegree centrality*) y el nodo que presenta un mayor número de enlaces salientes es el que presenta un mayor grado centralidad saliente (*outdegree centrality*).

La **intermediación** (*betweenness centrality*) de un nodo se relaciona con la capacidad de conectar o servir de puente entre nodos que no están directamente conectados. Se suele interpretar como una medida de la capacidad de un actor para influir en los flujos de información de una red y se considera que un actor con un alto grado de intermediación es relevante en la interconexión de diferentes subredes o clústers que puedan existir en la red. Normalmente, un nodo con un alto nivel de intermediación también presentará un alto nivel de centralidad.

Para calcular la intermediación en un grafo no dirigido, debemos sumar el número de veces que un nodo sirve de puente entre dos nodos no conectados directamente y normalizarlo dividiéndolo entre el número total de puentes entre parejas de nodos que se

pueden dar en la red. La herramienta VOSON ofrece un cálculo no normalizado de esta medida lo cual no consideramos un problema pues nos permite igualmente comparar los diferentes niveles de intermediación que presentan los nodos.

La **cercanía** (*closeness centrality*) de un nodo se relaciona con la distancia entre un actor y el resto de actores de la red y se suele interpretar como una medida de la capacidad que tiene un actor para interactuar con otros actores de la red. Para calcular esta medida debemos calcular el número mínimo de pasos que uniría a un nodo con el resto de nodos de la red, el número mínimo de pasos para unir nodo con otro es uno así que el número mínimo de pasos para conectar un nodo con todos los demás será el número total de nodos menos uno ($n-1$). Este valor deberemos dividirlo por el número real de pasos que tiene que dar en la red para conectarse con todos los nodos. VOSON ofrece un cálculo normalizado de esta medida que será el que utilizaremos en el análisis de los actores que forman parte de una red temática.

El algoritmo **HITS** (*Hypertext Induced Topic Selection*) es un algoritmo diseñado por Jon Kleinberg para valorar el prestigio de una página web. HITS usa dos indicadores, *HITS authority* y *HITS hub*, definiendo recursivamente cada uno a partir del otro. *HITS authority* valora la importancia de un nodo calculando la suma ponderada de valores *hub* de los enlaces que apuntan hacia ella. *HITS hub* determina la importancia de un nodo en función de los enlaces que dirige a otras páginas calculando la suma ponderada de los valores *HITS authority* de esas páginas a las que apunta. Como vemos el algoritmo HITS es una medida más compleja y refinada que las medidas de centralidad que hemos estudiado previamente y ha sido utilizada por los grandes buscadores para determinar el prestigio de las páginas que albergan en sus índices.

c) Visualización de redes de vecindario

Para analizar las relaciones de un actor con el resto de actores en una red utilizaremos el concepto de redes de vecindario. En este tipo de redes nos centramos en un único nodo y buscamos todas las relaciones directas con otros nodos considerando para ello los enlaces entrantes y salientes que presenta.

Para visualizar este tipo de redes de vecindario utilizaremos la opción de visualización de mapas “*Complete Network*” y seleccionaremos el nodo en que estamos interesados. La opción “*Highlight Nodes plus children*” nos permitirá visualizar con dos colores diferentes los nodos con enlaces entrantes o salientes. Como decíamos previamente, este análisis de las redes de vecindario se hará considerando la red absoluta y no la red de coenlaces para obtener toda la información disponible de los nodos.

Otras opciones de visualización que serán relevantes para el análisis de actores en nuestra investigación son las de modificar la apariencia de los nodos en función del ámbito geográfico o del ámbito de actividad al que pertenecen, representar los enlaces como dirigidos o no dirigidos, añadir a cada nodo el nombre del sitio o la URL o modificar el tamaño de los nodos en función del número de enlaces entrantes, salientes, etc. Todas ellas nos permitirán visualizar de manera gráfica las diferentes características de los actores que forman parte de la red.

Fases y Componentes del Análisis de Actores

Identificación de actores

- Agrupamiento
- Desagrupamiento

Análisis de redes sociales

- Centralidad
- Intermediación
- Cercanía
- HITS

Visualización de redes de vecindario

4.1.3. Fases y componentes para el análisis de temas

Una vez planteada la metodología que nos permite analizar las redes de hiperenlaces y los actores que forman parte de una red temática, en la tercera fase del análisis debemos

ser capaces de identificar las diferentes posiciones, subtemas o marcos de referencia (*frames*) que defienden los actores de la red.

Existen diferentes metodologías para analizar los subtemas o marcos de referencia que aparecen en una red temática y desarrollar representaciones gráficas que permiten visualizar la organización de la información. En primer lugar es imprescindible llevar a cabo un análisis del contenido de las páginas que van a formar el corpus que puede llevarse a cabo a través de un equipo humano que analice manualmente los documentos o compilando un corpus a través de la utilización de un *crawler* que sea capaz de filtrar los sitios web en función de las palabras clave que las caracterizan o incluso en función de los textos completos que aparecen en esos sitios web. En este sentido existen varios proyectos para obtener ontologías por medio de herramientas automáticas³⁴ que permitirían reducir notablemente el trabajo de este tipo de investigaciones enfocadas al estudio de los procesos de formación de la opinión pública en la Web.

Una vez identificada la red temática en cuestión y las subredes que aparecen en ella, se deberán extraer los textos que aparecen en los sitios web seleccionados. Para llevar a cabo esta tarea propongo utilizar SocSciBot, un programa que como hemos explicado previamente, permite extraer automáticamente los textos de un corpus de sitios web.

Después de realizar los análisis necesarios y extraer la información relevante del corpus de sitios web, el siguiente paso es construir el mapa temático a través del cual se organizará la información. Para hacer esto existen diferentes modelos de representación del conocimiento y lenguajes controlados.

Los lenguajes controlados son mecanismos utilizados para la presentación y organización del conocimiento que tienen por objetivo controlar y normalizar la asignación de palabras clave a un documento (Vallez et al., 2010). A continuación presentamos brevemente los modelos de lenguaje controlado que consideramos más relevantes para nuestra investigación:

34 CYC Ontology Project: <http://www.cyc.com>

- Taxonomías: Son una forma de clasificación jerárquica del conocimiento que permite relacionar unidades de información organizándolas en clases y subclases (Chris, 2007 citado en Valle et al., 2010).
- Tesoros: Son listas de términos usadas para organizar el conocimiento de un ámbito y controlar la descripción temática de un documento. Se basa en términos, descriptores y relaciones semánticas y funcionales que se establecen entre los términos. Estas relaciones pueden ser de equivalencia, asociación o jerarquía (López-Huertas, 99 citado en Valle et al., 2010).
- Ontologías: son la especificación del conocimiento de un determinado ámbito a través de la organización de conceptos, normalmente estructurados jerárquicamente, para representar entidades, ideas o eventos además de su propiedades y relaciones en un sistema de categorías (Gruber, 2008 citado en Nazar, en prensa).

Estas son las propuestas de lenguaje controlado que consideramos más importantes, pero existen otras propuestas para la representación del conocimiento entre las que destacan las redes semánticas de Quillian (1968), los gráficos conceptuales de Sowa (1991), el análisis conceptual formal (Ganter y Wille, 1999), el marco de los espacios conceptuales (Gärdenfors y Williams, 2001), los mapas conceptuales (Novak y Cañas, 2006) o los mapas temáticos (Park y Hunting, 2003).

Como señalábamos en el Capítulo 2, el término de red temática con el que trabajamos en esta investigación proviene del término en inglés *issue network*. En castellano, a diferencia del inglés, no existen dos términos para diferenciar lo que en inglés se define como *topic* e *issue*. Mientras se considera que un *topic* es simplemente una forma de generalización simbólica que posibilita la comunicación, se considera que un *issue* se forma a través de la defensa de posturas diferenciadas en una controversia generada en torno a un tema que podríamos definir como político. Por tanto, los temas políticos o *issues* son objetos de discusión que se desarrollan a través de la interacción entre actores sociales y políticos con distintas posiciones.

A su vez, como señalábamos en el Capítulo 3 sobre las Herramientas para el Análisis de Redes Temáticas, es posible identificar subtemas, marcos de referencia u opiniones en una red a partir del análisis de las informaciones de contexto relacionadas con uno o varios sitios. Los principales tipos de información de contexto son el contexto del enlace, las páginas enlazadas o el gráfico web; componentes que consideramos a continuación.

- El **contexto del enlace**: se refiere al contenido léxico que se encuentra relacionado con el enlace; este puede limitarse al texto del enlace o a todo el contenido de la página que contiene el enlace.
- Utilizar las páginas enlazadas como información de contexto supone tener en cuenta el **contenido léxico del texto** de los enlaces que recibe una página o de las páginas que envían sus enlaces hacia ella.
- El **gráfico web** como información de contexto supone analizar el gráfico que se forma a partir de los enlaces entrantes y salientes de una página sabiendo que las páginas relacionadas con un tema tienden a enlazarse más con páginas relacionadas con el mismo tema (Chakrabarti et al. 1999; Cho et al. 1998).

a) Análisis de comunidades

La metodología que proponemos para analizar los subtemas y marcos de referencia de una red temática tiene en cuenta tanto las comunidades como los actores agrupando los resultados en base a estas dos categorías.

El agrupamiento en clústers se ha realizado previamente en la fase del análisis de redes de hiperenlaces. Teniendo en cuenta las investigaciones de Chakrabarti et al. (1999) y Cho et al. (1998) podemos afirmar que los sitios que han sido agrupados en clústers comparten marcos de referencia comunes. El posterior análisis del contenido de los sitios que forman parte de cada clúster nos permitirá describir y caracterizar el marco de referencia asociado.

El análisis de contenido puede realizarse a través de métodos supervisados de aprendizaje (*supervised learning*) como Naïve Bayes, frecuencias de palabras (TD-IDF) o

espacios vectoriales. En nuestra metodología proponemos llevar a cabo un análisis de frecuencia de palabras ya que las herramientas estudiadas no ofrecen la posibilidad de realizar otro tipo de análisis más sofisticados.

VOSON ofrece la posibilidad de recuperar las palabras clave y las palabras y parejas de palabras que aparecen en el contenido de las páginas pero únicamente de las páginas que han sido introducidas como semillas en el rastreo. En el análisis de redes temáticas proponemos analizar los marcos de referencia que utilizan las diferentes comunidades que hemos identificado en la fase previa y para ello será necesario analizar el contenido de todas las páginas que forman parte de la comunidad y no únicamente de las semillas. De las herramientas evaluadas, consideramos que SocSciBot es la más adecuada ya que cuenta con el módulo Cyclist que permite indexar y llevar a cabo análisis de frecuencia de palabras en las páginas recuperadas.

Para llevar a cabo este tipo de análisis de contenido será necesario un procesamiento previo de la base de datos eliminando palabras vacías (*stop words*) y términos relacionados con el código de programación de los sitios web como *widget*, *hover* o *host* que suelen aparecer en la indexación de los contenidos de las páginas. En ocasiones podemos encontrar sitios que no incluyen palabras clave en su código fuente o páginas de las que no es posible extraer el contenido textual porque utilizan lenguajes de programación como Java que el crawler no puede interpretar.

Una vez depurada la información extraída de las páginas web procederemos a analizar su contenido. Para ello llevaremos a cabo un análisis de frecuencia de palabras en la página de inicio de los sitios de cada actor. Las páginas de inicio de los sitios web suelen ofrecer una descripción de la misión, visión y objetivos de la entidad a la que representa y un resumen de los contenidos del resto de páginas que forman parte del sitio. Por tanto, analizando únicamente el contenido de estas páginas iniciales podemos obtener una representación confiable del marco de referencia que defiende cada comunidad en la red.

b) Análisis de actores

Después de analizar los marcos de referencia de las comunidades de la red llevaremos a cabo el mismo tipo de análisis sobre los actores que se pueden considerar más

importantes de la red. Para definir cuáles son los actores más importantes de la red crearemos un ranking en base al número de enlaces entrantes que presenta cada actor.

Recordemos que los enlaces entrantes son una de las varias formas de medir la centralidad de un actor y en el caso del análisis de redes temáticas consideramos que esta medida es la que mejor representa la capacidad de influencia de un actor en la red. Al analizar los contenidos relacionados con un único actor pretendemos identificar con más detalle el marco de referencia que defiende ese actor y comprobar si coincide con el marco de referencia de la comunidad a la que pertenece o si presenta diferencias importantes con ella. Para conseguir un análisis más detallado proponemos analizar no únicamente las páginas de inicio del sitio web del actor, sino del sitio web completo.

c) Visualización de nubes de palabras

Los resultados del análisis se visualizarán en forma de nube de palabras asociados a las representaciones gráficas de los clústers y de las redes de vecindario de cada actor. El formato de visualización de nubes de palabras ha sido ampliamente utilizado en el análisis de contenidos y permite representar de una forma fácilmente comprensible los términos más frecuentes de un texto. Para crear las representaciones de nubes de palabras existen numerosas herramientas pero por su simplicidad no consideramos necesario caracterizarlas o evaluarlas en detalle. En nuestro caso utilizaremos Tag Crowd pero cualquier otra herramienta del mismo tipo puede resultar igualmente adecuada.

Fases y Componentes del Análisis de Temas.

Análisis de comunidades

- Frecuencia de palabras en las páginas de inicio del sitio

Análisis de actores

- Frecuencia de palabras en todas las páginas del sitio

Visualización de nubes de palabras

- Palabras asociadas a las comunidades
- Palabras asociadas a los actores

4.2. Estudio de caso: Análisis de la Red Temática de la Transparencia en Chile

En la primera parte de este Capítulo, hemos planteado una metodología que se divide en tres fases: análisis de redes de hiperenlaces, análisis de actores y análisis de temas. En esta segunda parte aplicaremos la metodología propuesta a un caso concreto que denominaremos la **Red Temática de la Transparencia en Chile**³⁵.

La selección de esta red se debe a que cumple con la definición de tema que hemos propuesto en capítulos anteriores, es decir, es un conjunto de ideas, teorías o conceptos que generan controversia entre grupos de actores políticos, sociales o económicos.

En este caso concreto hemos elegido el tema de la transparencia porque cumple ciertos requerimientos que lo hace idóneo para un trabajo de tipo exploratorio como este:

- En primer lugar, por supuesto, es un tema de debate o de discusión real. Es decir, de facto, existe en Chile un número de actores con opiniones distintas y que son activos en la web alrededor de este tema.
- En segundo lugar, hemos elegido deliberadamente un tema de debate sobre el cual en principio parece existir un consenso en relación a la conveniencia de que los gobiernos sean transparentes, es decir, que se sometan a una rendición de cuentas. Esto evita que la elección misma de tema de debate sea controvertido.
- En tercer lugar, el motivo de restringir el tema de discusión elegido a un ámbito geográfico concreto se debe a la necesidad de aplicar y probar la metodología propuesta sobre un tema de discusión conocido por el investigador, lo que consideramos aporta cierto grado de fiabilidad a la evaluación de los resultados siendo posible detectar irregularidades que podrían pasar desapercibidas si se analizasen temas de discusión completamente desconocidos.

³⁵ Entendemos la transparencia como la voluntad de saber y el derecho a acceder a la información de los ciudadanos.

Por último, una advertencia que quizá a estas alturas sea redundante pero que consideramos necesario enfatizar. Las Redes Temáticas no siempre son, o no son necesariamente, una realidad formal que se constituye como tal de forma expresa y explícita. Con frecuencia, son realidades implícitas que el análisis es capaz de identificar, definir y caracterizar.

Es el caso de la que nosotros hemos denominado Red Temática de la Transparencia en Chile. Partiendo de nuestro conocimiento de la realidad social y política chilena, planteamos la hipótesis (poco arriesgada, ciertamente) de que existía tal Red puesto que es bien conocido este debate. Un debate por cierto que seguramente se produce en otros muchos países con al menos un razonable nivel de democracia y ciertos grados de libertades.

A continuación, aplicaremos nuestro sistema de análisis a los distintos componentes, que según nuestra propuesta de modelo de Red Temática, constituye una de tales redes: los hiperenlaces, los temas y los actores.³⁶

4.2.1. Análisis de la red de hiperenlaces

Recordemos que la metodología propuesta para analizar redes de hiperenlaces se divide en tres fases: (1) selección de semillas, (2) rastreo de enlaces (3) análisis y visualización de redes.

a) Selección de semillas

Como señalábamos previamente, el método más habitual para desarrollar un rastreo preferencial definido por un tema comienza por seleccionar un grupo de sitios web relevantes en el área temática que queremos analizar. Para seleccionar ese grupo de sitios web relevantes que actuarán como semillas en el rastreo de redes de hiperenlaces proponíamos varias estrategias que se podían combinar en función de las necesidades o recursos del investigador. En nuestro caso hemos utilizado una estrategia que combina la

³⁶ Los datos obtenidos en las diferentes fases del análisis se pueden consultar en <https://drive.google.com/folderview?id=0B0IEptU7YxsXa0wwd0VLVGkyRFU&usp=sharing>

consulta en buscadores y la consulta de directorios. Consideramos que la combinación de estas dos estrategias nos permite evitar las limitaciones que presentan por separado y obtener así un grupo de semillas que sea representativo de los actores que desde diferentes ámbitos defienden posiciones en el tema de discusión relacionado con la transparencia en Chile.

Las **consultas en buscadores** se han hecho sobre Bing y Google utilizando las palabras clave que consideramos más relacionadas con el tema de discusión: transparencia y probidad. Para obtener resultados del ámbito geográfico de Chile hemos utilizado la opción que ofrecen estos dos buscadores de restringir los resultados por región y para evitar resultados personalizados en base a nuestro historial de búsquedas, ubicación geográfica, etc., hemos utilizado un navegador sin historial en exclusiva para la selección de semillas.

De las cuatro consultas realizadas se han recuperado los diez primeros resultados que hemos reducido a la URL de la página de inicio excepto en el caso de entidades cuya principal misión no es la transparencia. En estos casos se ha utilizado la URL del directorio, por ejemplo uchile.cl/transparencia o mineduc.cl/transparencia que representan las secciones de la Universidad de Chile y del Ministerio de Educación dedicadas a la transparencia. En total se han obtenido 18 resultados relevantes que ofrecemos a continuación en la Tabla 12.

www.gobiernotransparentechile.cl	www.agendadeprobidad.gob.cl
www.consejotransparencia.cl	www.cdc.gob.cl
www.portaltransparencia.cl	www.subtrans.gob.cl/transparencia
www.mineduc.cl/transparencia	transparencia.redsalud.gob.cl
www.educatransparencia.cl	www.transparenciamaipu.cl
www.leydetransparencia.cl	www.leychile.cl
www.uchile.cl/transparencia	transparencia.dga.cl
www.dt.gob.cl/transparencia	www.datos.gob.cl
www.probidadenchile.cl	

Tabla 12. Semillas obtenidas a través de consultas en buscadores. Elaboración propia.

Las semillas obtenidas a través de la consulta de buscadores se han ampliado a través de la **consulta de los directorios** que mantienen la Sunlight Foundation³⁷, Opening Parliament³⁸ y Open Government Partnership³⁹. A través de estos tres directorios se han identificado seis entidades de tipo no gubernamental que no habían aparecido en las consultas a los buscadores donde se habían obtenido mayoritariamente resultados del ámbito gubernamental. Como señalábamos en la propuesta de metodología, en la selección del grupo de semillas es necesario mantener un equilibrio entre organizaciones de diferentes ámbitos que puedan intervenir en el tema de discusión. A continuación mostramos el conjunto total de semillas en la Tabla 13.

www.gobiernotransparentechile.cl	www.agendadeprobidad.gob.cl
www.consejotransparencia.cl	www.cdc.gob.cl
www.portaltransparencia.cl	www.subtrans.gob.cl/transparencia
www.mineduc.cl/transparencia	transparencia.redsalud.gob.cl
www.educatransparencia.cl	www.transparenciamaipu.cl
www.leydetransparencia.cl	www.leychile.cl
www.uchile.cl/transparencia	transparencia.dga.cl
www.dt.gob.cl/transparencia	www.datos.gob.cl
www.probidadenchile.cl	www.ciudadanointeligente.org
www.proacceso.cl	www.chiletransparente.cl
www.probono.cl	www.poderopedia.org
www.ciudadviva.cl	

Tabla 13. Semillas obtenidas a través de consultas en buscadores y directorios. Elaboración propia.

³⁷ Sunlight Foundation The Sunlight Foundation is a national, nonpartisan, nonprofit organization that uses the tools of civic tech, open data, policy analysis and journalism to make our government and politics more accountable and transparent to all. <http://sunlightfoundation.com/>

³⁸ OpeningParliament.org is a forum intended to help connect the world's civic organizations engaged in monitoring, supporting and opening up their countries' parliaments and legislative institutions. It also serves as the home of the Declaration on Parliamentary Openness, a set of shared principles on the openness, transparency and accessibility of parliaments supported by more than 140 organizations from over 75 countries. <http://www.openingparliament.org/>

³⁹ The Open Government Partnership was formally launched on September 20, 2011, when the 8 founding governments (Brazil, Indonesia, Mexico, Norway, Philippines, South Africa, United Kingdom, United States) endorsed the Open Government Declaration, and presented their country action plans. Since September 2011, OGP has welcomed the commitment of more than 50 additional governments to join the Partnership. www.ogphub.org

b) Rastreo de enlaces

Una vez seleccionado un grupo de semillas que se puede considerar representativo de los actores que participan en el tema de discusión desde diferentes ámbitos, hemos procedido a desarrollar un rastreo de los enlaces que presentaban las semillas. Para ello utilizamos la herramienta VOSON que como dijimos en el Capítulo 3, es la única que ofrece la posibilidad de llevar a cabo simultáneamente un análisis de enlaces entrantes y salientes.

El rastreo de los enlaces entrantes se configuró para detenerse después recuperar un máximo de 1000 enlaces entrantes por sitio (*Crawl inbound*) detectando para ello los enlaces entrantes que se producían sobre páginas internas de cada semilla (*Internal inbound links*).

El rastreo de los enlaces salientes se configuró para detenerse tras detectar un máximo de 1000 enlaces salientes por semilla (*Crawl outbound*), después de detectar 25 páginas internas de cada semilla que no contuviesen enlaces salientes (*Max. Unproductive pages*), después de rastrear 100 páginas en cada semilla (*Depth of crawl pages*) en un máximo de tres niveles de profundidad (*Depth of crawl levels*).

Además de estas opciones de configuración relacionadas con el rastreo de enlaces, también se eligió la opción de recuperar los textos contenidos en las semillas para posteriormente llevar a cabo un análisis de contenido que nos permita identificar las diferentes posiciones defendidas por los actores en relación al tema de discusión. Esto lo veremos con más detalle en el apartado 4.1.3. Análisis de temas.

El análisis de enlaces entrantes y salientes del grupo de semillas ha dado como resultado 26.250 URLs⁴⁰. Estas URLs son agrupadas automáticamente por VOSON en base a la

⁴⁰ Los resultados del análisis de enlaces entrantes y salientes de la semillas se adjuntan en formato CSV en el Anexo X de esta investigación con el nombre Red Temática Absoluta de la Transparencia en Chile

coincidencia de su dominio, dando como resultado 1.874 grupos de páginas⁴¹ que representan los sitios de los actores que forman parte de la red.

Como señalábamos en la primera parte de este Capítulo 4, el agrupamiento automático de VOSON puede presentar inexactitudes. En caso de detectar algún sitio que no ha sido agrupado correctamente utilizaremos la opción *pagegrouping* o *preserving* para agrupar o separar páginas. En nuestro caso, en una revisión superficial de los resultados hemos detectado este tipo de casos pero no consideramos que modifiquen sustancialmente los resultados por lo que no hemos realizado ningún proceso de agrupamiento adicional.

Este tipo de depuración requiere en ocasiones un análisis cualitativo manual de las páginas y eso es precisamente lo que queremos evitar con nuestra metodología para el análisis de redes temáticas. Por lo tanto, en este caso hemos descartado este tipo de depuración manual de los resultados.

Entre los resultados obtenidos podremos encontrar sitios no relevantes relacionados con aspectos técnicos de los sitios como creativecommons.org o w3c.validator.org que no tienen relación con la red temática a pesar de ser enlazados por el grupo de semillas.

Para eliminar este tipo de sitios podemos utilizar la opción *pruning* que ofrece VOSON en la que simplemente introduciremos las URL de los sitios que no queremos considerar en la red. Igual que sucede con el proceso de agrupamiento manual, la eliminación de sitios no relevantes requiere de un análisis cualitativo manual de los resultados lo que requiere de grandes recursos de tiempo. Tras una revisión superficial hemos podido comprobar que el número de resultados no relevantes es relativamente pequeño y en ningún caso modifican la estructura de la red. Por lo tanto, hemos descartado llevar a cabo este tipo de depuración.

Una vez identificados los enlaces de la red y los actores que forman parte de ella hemos procedido a refinar la red mediante el análisis de coenlaces y la identificación de enlaces recíprocos entre el grupo de semillas y el resto de nodos de la red. Como indicábamos en apartados anteriores, el análisis de coenlaces nos permite identificar los actores que

⁴¹ Los resultados del agrupamiento de URLs se adjuntan en formato CSV en el Anexo X con el nombre Red Temática Agrupada de la Transparencia en Chile.

tienen un mayor grado de similitud con las semillas y que por tanto se pueden considerar más relacionados con el tema de discusión. La identificación de enlaces recíprocos entre los sitios del grupo de semillas y el resto de nodos de la red nos permite detectar sitios que presentan relaciones de colaboración o pertenencia con el grupo de semillas y que por tanto también pueden ser considerados en relación con el tema de discusión.

La depuración y eliminación de los nodos menos relevantes nos permite facilitar la interpretación de la red. Para llevar a cabo el análisis de coenlaces y la identificación de enlaces recíprocos utilizamos la opción “*create seeds y seeds + important sites*” que incorpora VOSON. El análisis de coenlaces y enlaces recíprocos ha dado como resultado una red formada por 474 nodos frente a los 1.874 de la red absoluta obtenida inicialmente.

c) Análisis y Visualización

En nuestra metodología proponíamos analizar la red absoluta y la red refinada de coenlaces y enlaces recíprocos a través del cálculo de la densidad (*density*), la centralidad (*degree centrality*), la intermediación (*betweenness centrality*) y la cercanía (*closeness centrality*). Y posteriormente identificar los *clústers* o comunidades que existen en la red de coenlaces y analizar nuevamente las medidas señaladas para cada uno de ellos.

En la Figura 60 presentamos la visualización de la Red Temática de la Transparencia en Chile en su versión absoluta o no refinada, es decir, considerando todos los nodos y enlaces entrantes y salientes existentes.

En la Figura 61 presentamos la visualización de la Red Temática de la Transparencia en Chile en su versión depurada y diferenciando los clústeres de la red con diferentes colores. Es decir, se han considerado solo los nodos que presentan enlaces recíprocos o coenlaces con el grupo de semillas y posteriormente se ha desarrollado un agrupamiento clustering que se representa a través de diferentes colores.

En el caso de la red absoluta y la red refinada, las visualizaciones se han hecho aplicando el algoritmo Fruchterman-Reingold con 100 iteraciones, condicionando el tamaño de los

nodos al número de enlaces entrantes y representando los enlaces como no dirigidos. En el caso de estas dos redes hemos añadido los nombres de los sitios web más importantes con el objetivo de aportar cierto significado a las representaciones.

Los resultados del análisis de redes sociales son los esperados si tenemos en cuenta que la red absoluta considera los enlaces entrantes y salientes sin otro criterio de discriminación. La red refinada considera igualmente enlaces entrantes y salientes pero restringe los resultados a aquellos nodos que presentan un coenlace o un enlace recíproco con el grupo de semillas. El número de nodos se reduce de 1874 a 474 y el número de enlaces de 2606 a 1206. La densidad aumenta en un orden de diez unidades y la centralidad, la intermediación y la cercanía se multiplican entre dos y tres veces. Todo lo cual nos habla de una red en que las relaciones entre los actores son más estrechas y dinámicas.

El análisis comparado de redes sociales nos permite identificar algunas particularidades de su funcionamiento. En este caso estamos comparando dos versiones, absoluta y refinada, de una misma red y los resultados, confirman lo esperado. Probablemente, este tipo de análisis ofrezca una información de mayor valor en la comparación de las redes temáticas de la transparencia de varios países, lo que esperamos poder comprobar en investigaciones futuras.

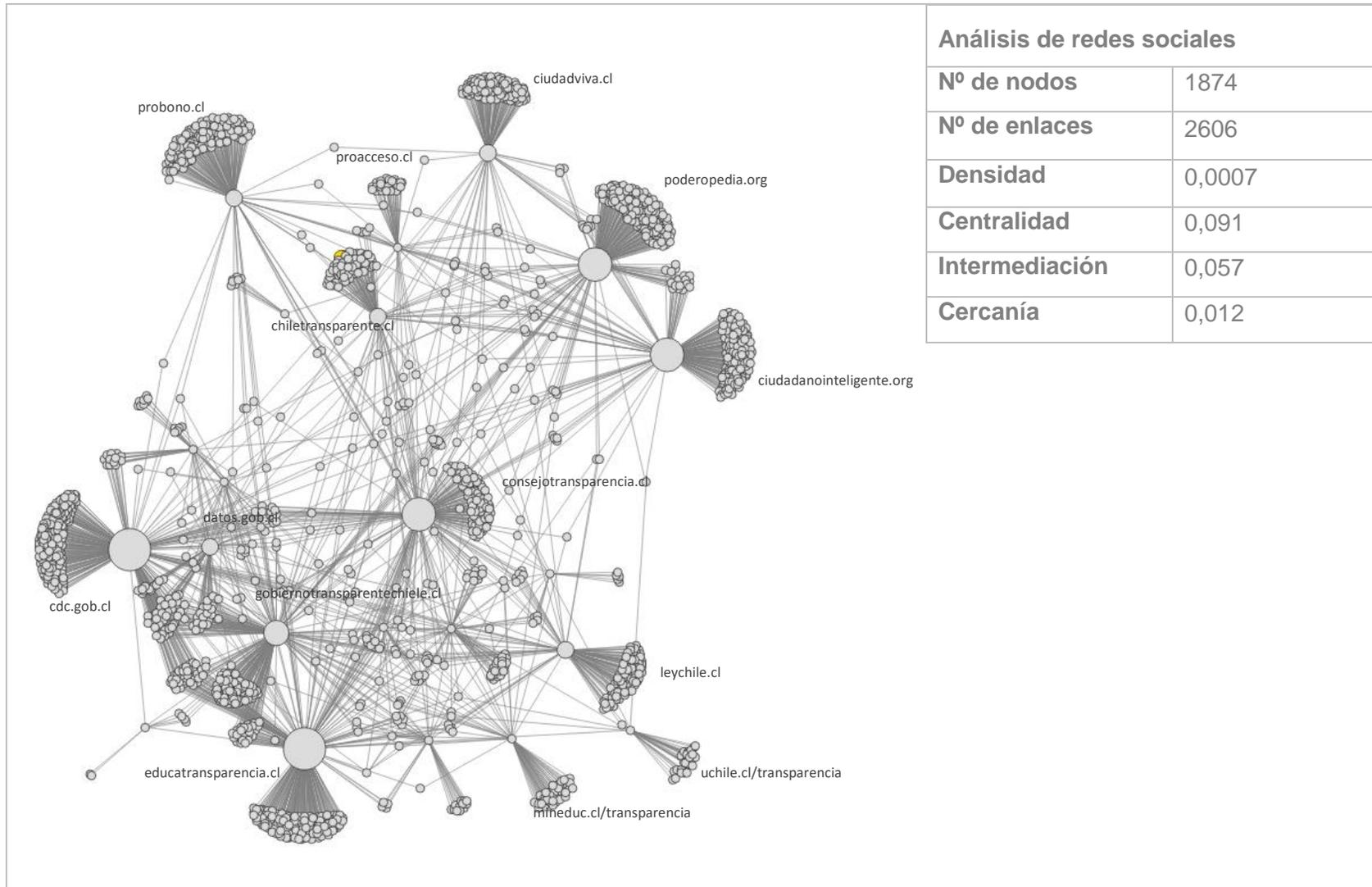


Figura 60. Red Temática absoluta de la Transparencia en Chile

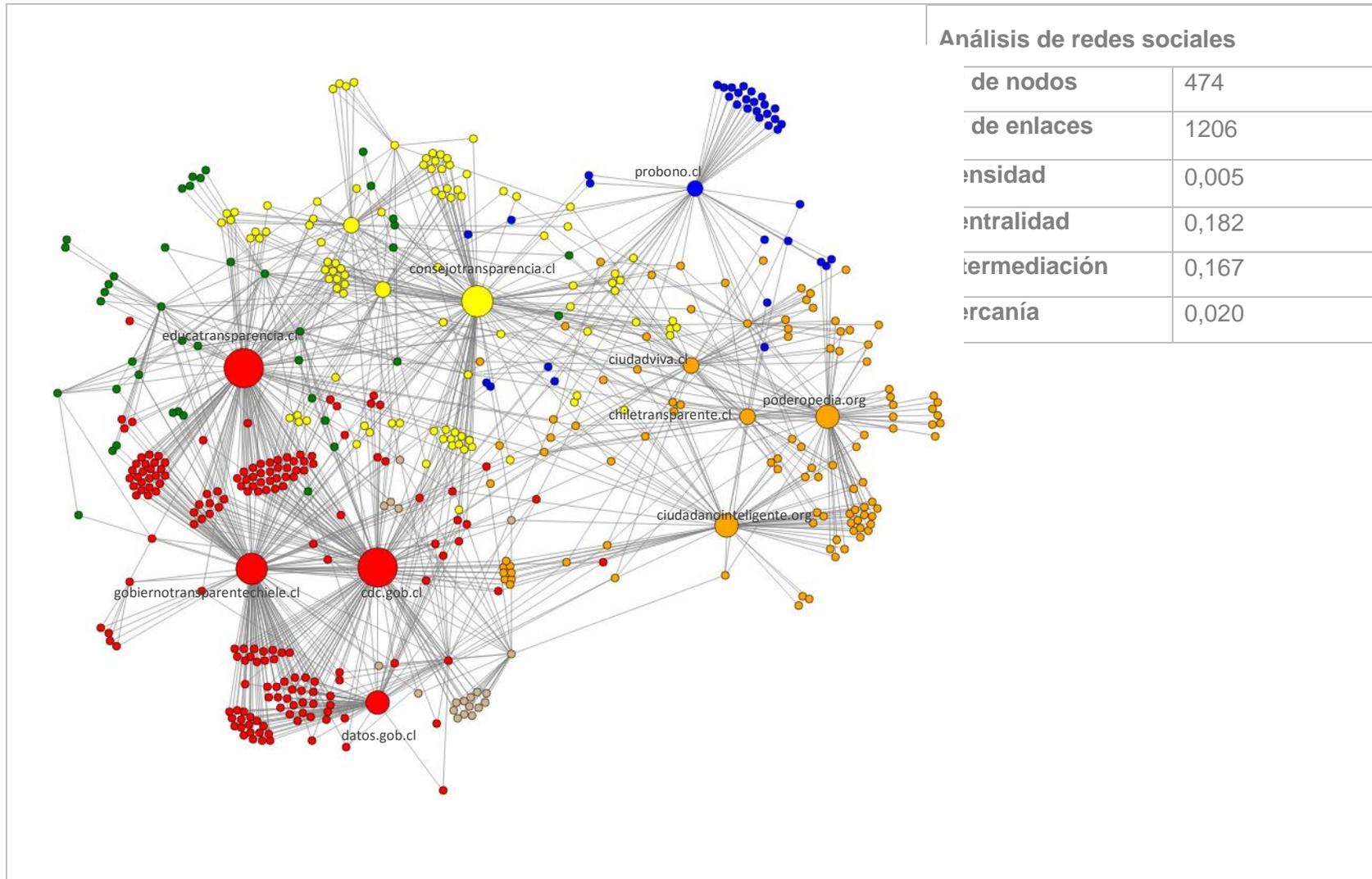


Figura 61. Red Temática de la Transparencia en Chile refinada considerando únicamente enlaces recíprocos y coenlaces y dividida en comunidades por colores

Como indicábamos en la primera parte de este capítulo, el agrupamiento en *clústeres* o *clustering* permite la identificación de comunidades dentro de una red. Para identificar estas comunidades VOSON aplica el algoritmo Newman-Girvan que utiliza el concepto de intermediación (*betweenness*) aplicado a los enlaces de la red (*edges betweenness*). A través del cálculo y eliminación de los enlaces que presentan mayor nivel de intermediación será posible identificar los grupos de nodos desconectados que se pueden considerar *clústeres*.

Al aplicar este algoritmo en la Red Temática de la Transparencia en Chile se han identificado 6 *clústeres* o comunidades que han sido convertidos en subredes para poder calcular las medidas de densidad, centralidad, intermediación y cercanía en cada una de ellas.

Para convertir en subredes las comunidades que hemos identificado previamente con la aplicación del algoritmo Newman-Girvan, es decir, para convertirlas en redes independientes, hemos empleado la utilidad "*composition*" de VOSON que nos permite dividir la red en función de diferentes criterios; en este caso, el criterio empleado en la creación de nuevas subredes ha sido el número de clúster que nos ha permitido crear una subred independiente para cada comunidad.

Recordemos que la utilidad de visualización de VOSON ofrece la posibilidad de interactuar con las imágenes y obtener información detallada de los nodos simplemente pasando el ratón sobre ellos. En este formato no es posible ofrecer la información asociada con cada representación gráfica pero consideramos que este tipo de información es fundamental para aportar el significado real que se obtiene con este sistema de análisis y con la herramienta seleccionada para llevarlo a cabo. Para tratar de suplir esta limitación hemos acompañado las representaciones actores con tablas en las que se identifican los actores más relevantes relacionados con cada clúster y/o actor. En el caso de la representación de las comunidades se han numerado los nodos para facilitar la identificación de los actores que las componen y las relaciones que establecen. Se ha añadido además las informaciones que consideramos más importantes sobre cada actor: nombre, número de páginas que contiene su sitio, número de enlaces entrantes, número de enlaces salientes, número de enlaces totales, nivel HITS Hub, nivel HITS Authority, nivel de cercanía y nivel de intermediación.

En la Tabla 14, presentamos las medidas de densidad, centralidad, intermediación y cercanía de cada uno de los clústeres. En este caso se ha aplicado un redondeo de tres decimales para facilitar la comprensión. En el caso de las tablas asociadas a las representaciones de clústers y actores hemos utilizado un redondeo de dos decimales para no sobrecargar de información las tablas y tratar de facilitar su lectura.

En relación a la densidad de los diferentes clústeres podemos señalar que todos ellos presentan una densidad baja o muy baja, alcanzando solo uno de ellos el nivel de 0.05. Los clústeres 4 y 6 son los que mayor grado de densidad presentan con un nivel de 0.043 y 0.053 respectivamente. Esto indica un alto nivel de dispersión en las redes lo que se puede interpretar como un bajo nivel de cohesión y relación entre los nodos que forman la red.

En relación al nivel de centralidad, todos los clústeres excepto nuevamente los clústeres número 4 y 6 presentan un grado de centralidad medio o bajo menor a 0.5. Esto muestra un bajo nivel de control en las redes por uno o pocos actores. En el caso del clúster 4 vemos exactamente lo contrario. El alto nivel de centralidad que alcanza un nivel de 0.824 indica que se trata de una red fuertemente dominada por un actor lo que podemos comprobar en la Figura 65.

En relación a la intermediación todos los clústeres presentan un nivel bajo o muy bajo excepto nuevamente el número 4. Esto se corresponde con redes en las que no existen nodos con un alto poder de control sobre las conexiones indirectas entre nodos y por tanto con un bajo poder de control sobre los flujos de información. Podemos comprobar en las Figuras 58 y 61 que representan a los clústeres 2 y 5, que se trata de redes en las que apenas existen rutas que permitan conectar nodos que no estén ya conectados directamente. La mayoría de los enlaces en estos clústeres se dirigen a un único nodo y este nodo no dirige enlaces hacia el resto de la red. Por tanto, es posible interpretar que los flujos de información en estos clústers son escasos.

En la representación del clúster 4 en la Figura 65 podemos comprobar al contrario que se trata de una red en la que el nodo central, el que mayor número de enlaces entrantes recibe, ofrece a su vez abundantes enlaces hacia el resto de nodos de la red lo que

permite el enlace indirecto entre nodos y los flujos de información entre nodos no conectados directamente.

En relación a la cercanía, nuevamente todos los nodos presentan bajos niveles excepto en el caso del Clúster 4. Esto se corresponde con redes en las que los nodos están poco conectados entre sí, siendo que la mayoría de los enlaces se dirigen hacia unos pocos nodos muy conectados mientras el resto no recibe enlaces. En el lado contrario, el Clúster 4 vuelve a presentar un alto grado de cercanía lo que se debe al gran número de enlaces salientes y recíprocos que existen en la red. Esto es indicativo de redes donde se producen relaciones fuertes de cooperación, alianza o incluso de pertenencia.

	Nº de nodos	Nº de Enlaces	Densidad	Centralidad	Intermediación	Cercanía
Clúster 1	171	404	0,014	0,370	0,261	0,046
Clúster 2	41	66	0,040	0,273	0,085	0,092
Clúster 3	104	212	0,020	0,321	0,195	0,020
Clúster 4	38	61	0,043	0,824	0,680	0,290
Clúster 5	100	151	0,015	0,407	0,126	0,019
Clúster 6	20	20	0,053	0,526	0,216	0,036

Tabla 14. Medidas de análisis de redes sociales de los clústers 1-6. Elaboración propia.

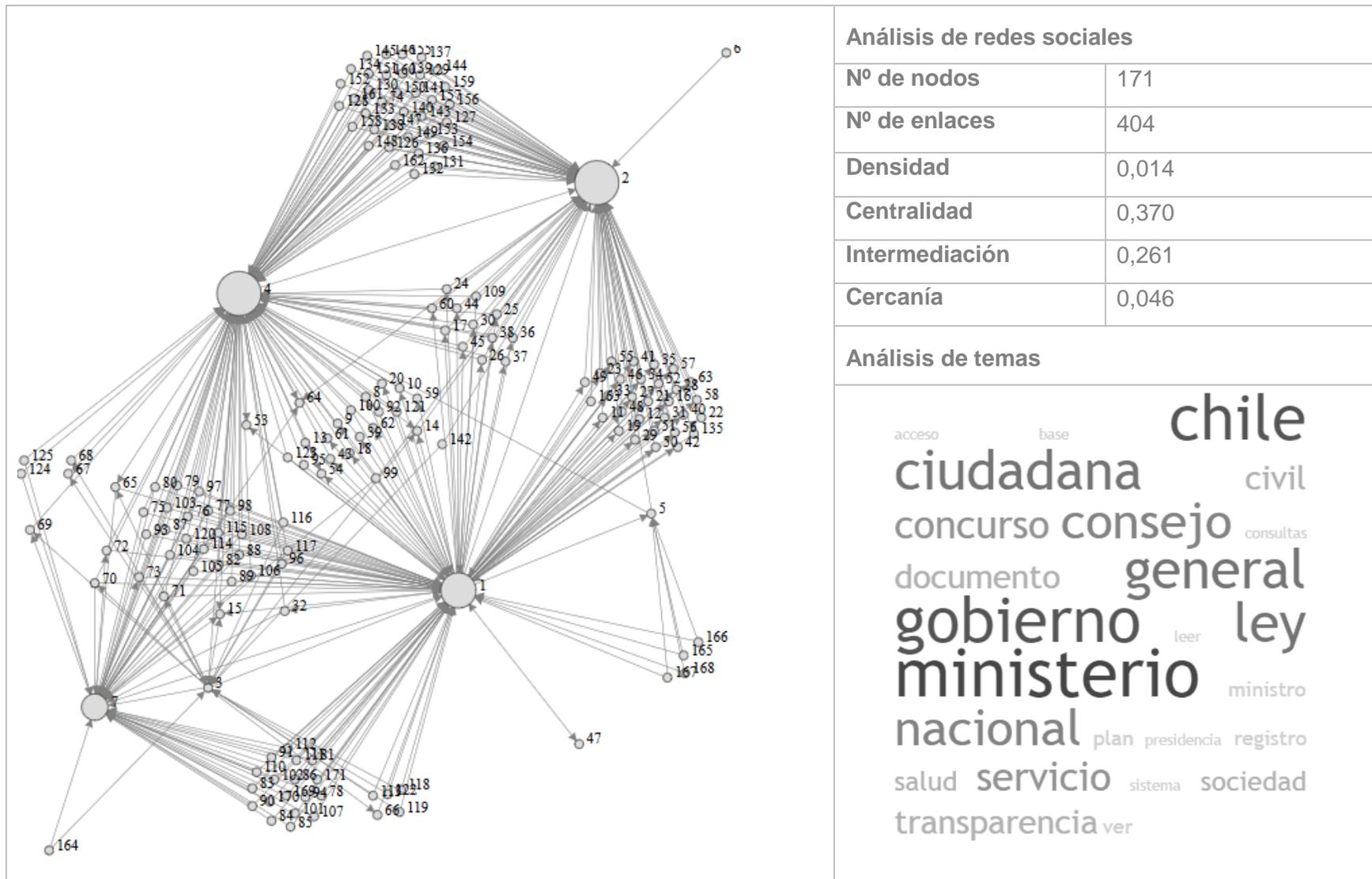


Figura 62. Representación gráfica y análisis de redes sociales y de temas del Clúster 1

Nº	Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
4	www.cdc.gob.cl	57	103	11	114	0,11	0,70	0,27	4851,54
2	www.educatransparencia.cl	134	85	1	86	0,00	0,43	0,20	62,50
1	www.gobiernotransparentechile.cl	1674	70	59	129	0,14	0,44	0,29	7549,85
7	www.datos.gob.cl	80	50	8	58	0,09	0,34	0,27	653,41
3	www.leydetransparencia.cl	62	12	12	24	0,12	0,06	0,27	639,03
5	www.subtrans.gob.cl	25	5	2	7	0,09	0,02	0,24	20,62
15	www.minsegpres.gob.cl	9	4	4	8	0,12	0,03	0,27	608,52
14	transparencia.minsegpres.gob.cl	6	3	1	4	0,03	0,03	0,20	0,53
53	www.gob.cl	44	2	1	3	0,05	0,02	0,24	0,00
65	participacionciudadana.minsegpres.gob.cl	32	2	3	5	0,11	0,02	0,27	0,00
67	www.opengovpartnership.org	5	2	0	2	0,00	0,02	0,20	0,00
69	www.comisiondefensoraciudadana.cl	1	2	1	3	0,05	0,02	0,24	0,00
70	www.modernizacion.gob.cl	25	2	3	5	0,11	0,02	0,27	0,00
8	www.ssffaa.gob.cl	2	1	1	2	0,05	0,01	0,24	0,00
9	www.ssffaa.cl	27	1	1	2	0,05	0,01	0,24	0,00
10	www.ind.cl	25	1	1	2	0,05	0,01	0,24	0,00
11	www.registrocivil.cl	10	1	1	2	0,03	0,01	0,20	0,00
12	transparenciaactiva.cultura.gob.cl	8	1	1	2	0,03	0,01	0,20	0,00
13	www.cultura.gob.cl	26	1	1	2	0,05	0,01	0,24	0,00
16	webhosting.redsalud.gov.cl	13	1	1	2	0,03	0,01	0,20	0,00
17	www.sec.cl	9	1	2	3	0,09	0,01	0,24	0,00
18	www.dsp.gob.cl	1	1	1	2	0,05	0,01	0,24	0,00
19	www.economia.cl	11	1	1	2	0,03	0,01	0,20	0,00
20	www.subtel.gob.cl	6	1	1	2	0,05	0,01	0,24	0,00
21	www.subtel.cl	2	1	1	2	0,03	0,01	0,20	0,00

Tabla 15. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 1

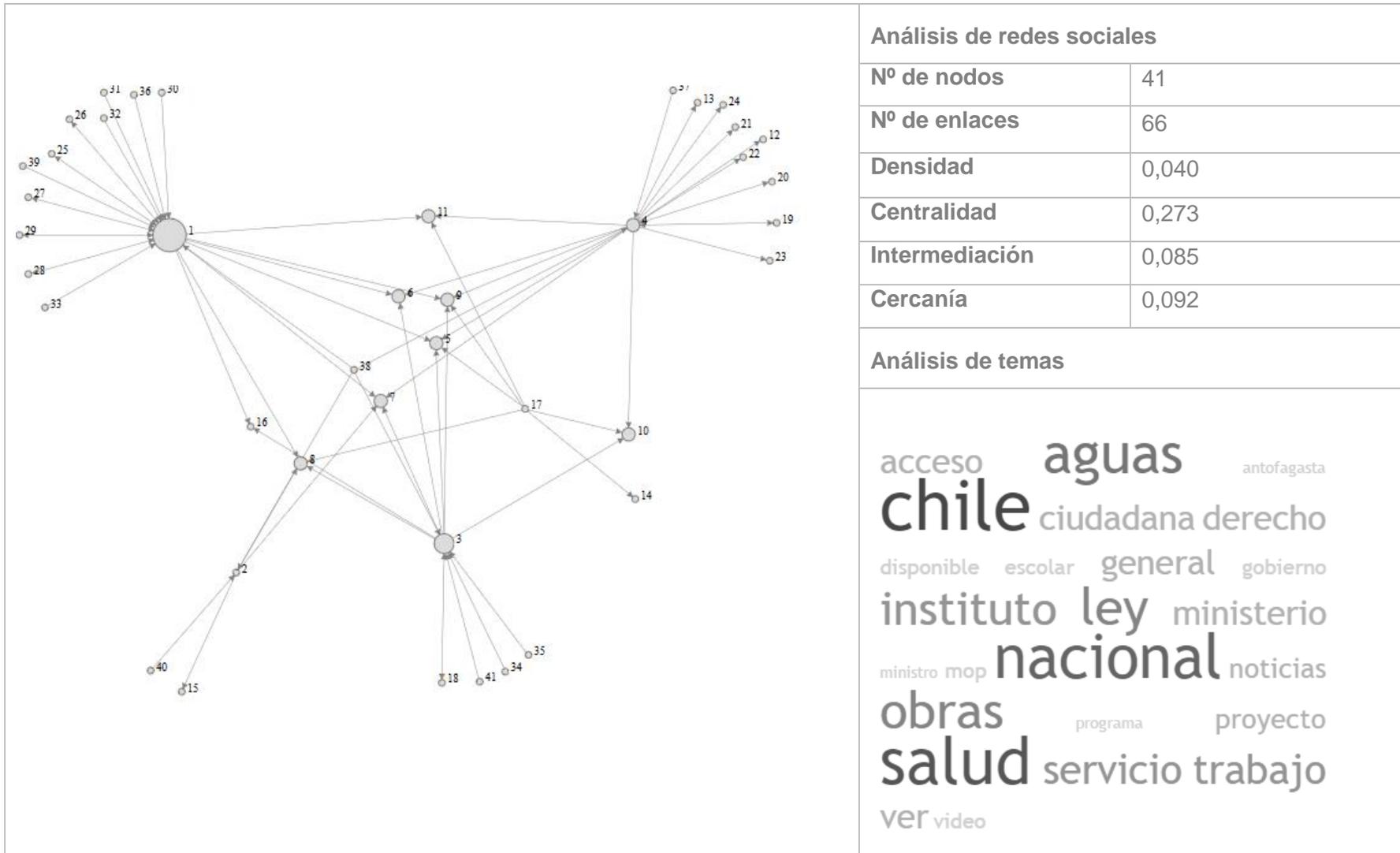


Figura 63. Representación gráfica y análisis de redes sociales y de temas del Clúster 2

Nº	Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
1	www.mineduc.cl	379	12	12	24	0,56	0,00	0,42	134,58
3	www.dt.gob.cl	9265	5	8	13	0,46	0,00	0,38	34,58
4	transparencia.dga.cl	301	4	14	18	0,59	0,00	0,43	50,25
5	www.dipres.gob.cl	65	4	0	4	0,00	0,39	0,33	0,00
7	www.mercadopublico.cl	59	4	0	4	0,00	0,35	0,33	0,00
8	www.registros19862.cl	35	4	0	4	0,00	0,30	0,33	0,00
9	www.gobiernotransparente.gob.cl	3	4	0	4	0,00	0,39	0,33	0,00
6	www.chileatiende.cl	47	3	0	3	0,00	0,32	0,33	0,00
10	www.gobiernotransparente.cl	35	3	0	3	0,00	0,28	0,33	0,00
11	www.gobiernotransparente.gov.cl	3	3	0	3	0,00	0,30	0,33	0,00
2	www.uchile.cl	7	2	3	5	0,13	0,00	0,35	4,58
16	dpi.minsegpres.gob.cl	3	2	0	2	0,00	0,20	0,33	0,00
12	www.dcyf.cl	2	1	0	1	0,00	0,12	0,33	0,00
13	www.mop.cl	13	1	0	1	0,00	0,12	0,33	0,00
14	www.minsal.cl	1	1	0	1	0,00	0,07	0,33	0,00
15	www.icei.uchile.cl	2	1	0	1	0,00	0,03	0,33	0,00
18	tramites.dirtrab.cl	11	1	1	2	0,00	0,09	0,36	0,00
19	www.dga.cl	29	1	1	2	0,00	0,12	0,38	0,00
20	snia.dga.cl	4	1	1	2	0,00	0,12	0,38	0,00
21	servicios.vialidad.cl	2	1	0	1	0,00	0,12	0,33	0,00
22	transparencia.arquitecturamop.cl	2	1	0	1	0,00	0,12	0,33	0,00
23	www.inh.cl	5	1	0	1	0,00	0,12	0,33	0,00
24	esp.aeropuertos.gov.cl	1	1	0	1	0,00	0,12	0,33	0,00
25	www.comunidadescolar.cl	31	1	1	2	0,00	0,11	0,37	0,00
26	www.mecsup.cl	77	1	1	2	0,00	0,11	0,37	0,00

Tabla 16. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 2

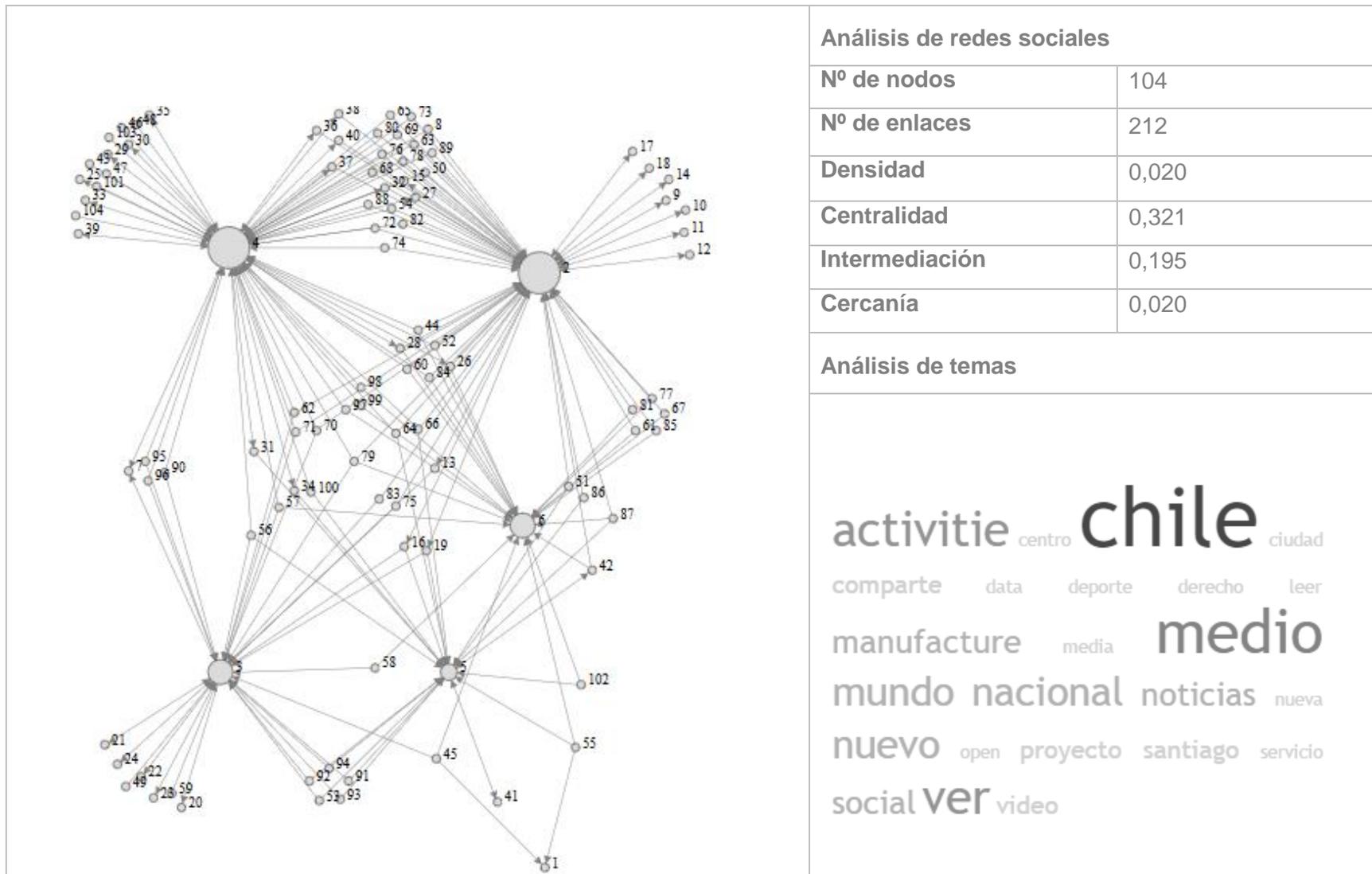


Figura 64. Representación gráfica y análisis de redes sociales y de temas del Clúster 3

Nº	Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
2	www.ciudadanointeligente.org	53	11	47	64	0,00	0,68	0,19	2097,00
4	www.poderopedia.org	52	17	443	69	0,00	0,65	0,19	1954,50
3	www.chiletransparente.cl	27	6	227	33	0,00	0,20	0,15	523,00
6	www.ciudadviva.cl	22	0	1	22	0,00	0,24	0,14	0,00
5	www.proacceso.cl	20	2	22	22	0,00	0,14	0,17	542,50
1	www.probidadenchile.cl	2	0	2	2	0,00	0,01	0,14	0,00
7	www.twitter.com	2	0	10	2	0,00	0,00	0,14	0,00
9	blog.ciudadanointeligente.org	1	1	7	2	0,07	0,00	0,18	0,00
10	poplus.org	1	1	13	2	0,07	0,00	0,18	0,00
11	hayacuerdo.ciudadanointeligente.org	1	1	2	2	0,07	0,00	0,18	0,00
12	www.deldichoalhecho.cl	1	1	5	2	0,07	0,00	0,18	0,00
13	www.transparencialegislativa.org	1	2	3	3	0,09	0,00	0,18	146,50
14	votainteligente.cl	1	1	10	2	0,07	0,00	0,18	0,00
16	quientefinancia.cl	1	2	3	3	0,04	0,00	0,17	200,75
17	www.candideit.org	1	1	25	2	0,07	0,00	0,18	0,00
18	congresoabierto.cl	1	1	25	2	0,07	0,00	0,18	0,00
19	leydelobby.cl	1	2	6	3	0,04	0,00	0,17	200,75
20	www.chilecompra.cl	1	0	2	1	0,00	0,00	0,14	0,00
21	www.transparency.org	1	1	10	2	0,02	0,00	0,15	0,00
22	www.partidostransparentes.cl	1	1	3	2	0,02	0,00	0,15	0,00
23	www.programamujer.cl	1	1	25	2	0,02	0,00	0,15	0,00
24	www.walmartchile.cl	1	1	2	2	0,02	0,00	0,15	0,00
25	www.cooperativa.cl	1	0	5	1	0,00	0,00	0,14	0,00
26	www.theclinic.cl	1	2	8	3	0,01	4E-32	0,18	71,63
27	radio.uchile.cl	1	1	5	2	0,07	3E-32	0,18	472,08

Tabla 17. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 3



Figura 65. Representación gráfica y análisis de redes sociales y de temas del Clúster 4

Nº	Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
1	www.probono.cl	409	31	30	61	0,03	0,69	0,73	906,00
2	www.sernam.cl	7	1	0	1	0,00	0,13	0,33	0,00
3	www.flickr.com	2	1	0	1	0,00	0,13	0,33	0,00
4	www.facebook.com	3	1	0	1	0,00	0,13		0,00
5	www.redprobono.org	26	1	1	2	0,18	0,13	0,33	0,00
6	www.az.cl	3	1	1	2	0,18	0,13	0,33	0,00
7	www.gqmc.cl	4	1	1	2	0,18	0,13	0,33	0,00
8	www.agycia.cl	3	1	1	2	0,18	0,13	0,33	0,00
9	www.guerrero.cl	3	1	1	2	0,18	0,13	0,33	0,00
10	www.fermandois.cl	26	1	1	2	0,18	0,13	0,33	0,00
11	www.palma.cl	25	1	1	2	0,18	0,13	0,33	0,00
12	www.alessandri.cl	4	1	1	2	0,18	0,13	0,33	0,00
13	www.rcz.cl	4	1	1	2	0,18	0,13	0,33	0,00
14	www.moralesybesa.cl	3	1	1	2	0,18	0,13	0,33	0,00
15	www.pfeffer.cl	3	1	1	2	0,18	0,13	0,33	0,00
16	www.bye.cl	2	1	1	2	0,18	0,13	0,33	0,00
17	www.gutierrezabogados.cl	18	1	1	2	0,18	0,13	0,33	0,00
18	www.bbis.cl	2	1	1	2	0,18	0,13	0,33	0,00
19	www.garciamagliona.cl	2	1	1	2	0,18	0,13	0,33	0,00
20	www.aranis-espinoza.cl	10	1	1	2	0,18	0,13	0,33	0,00
21	www.cdrabogados.cl	5	1	1	2	0,18	0,13	0,33	0,00
22	www.ayudar.cl	26	1	0	1	0,00	0,13	0,33	0,00
23	www.slideshare.net	2	1	1	2	0,18	0,13	0,33	0,00
24	www.horvitz.cl	6	1	1	2	0,18	0,13	0,33	0,00
25	www.vancecenter.org	6	1	1	2	0,18	0,13	0,33	0,00

Tabla 18. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 4

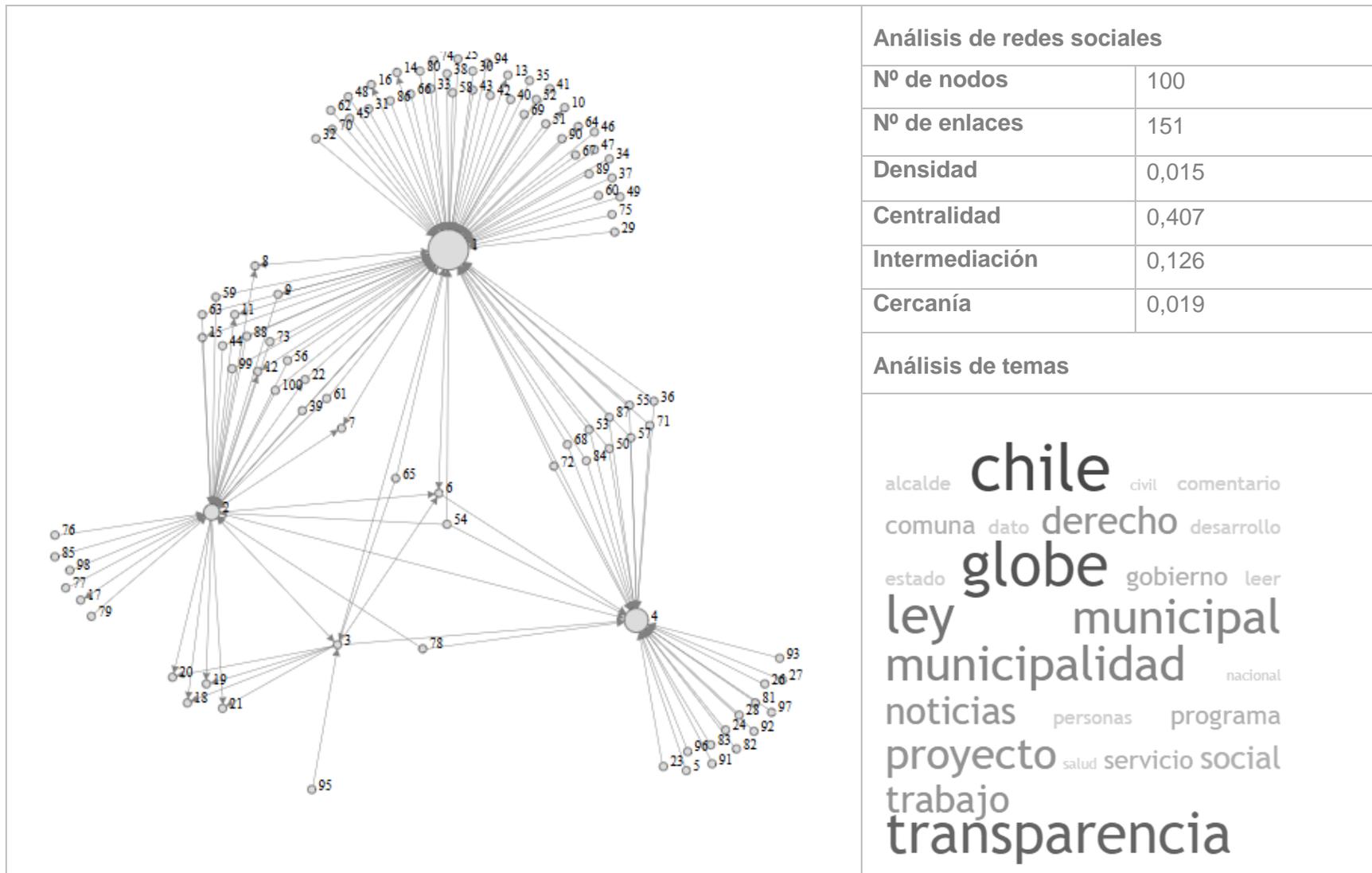


Figura 66. Representación gráfica y análisis de redes sociales y de temas del Clúster 5

Nº	Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
1	www.consejotransparencia.cl	432	69	13	82	0,09	0,91	0,29	1232,50
4	www.leychile.cl	1093	30	0	30	0,00	0,30	0,25	0,00
2	www.portaltransparencia.cl	493	22	14	36	0,18	0,26	0,29	654,00
3	www.transparenciamaipu.cl	184	3	8	11	0,19	0,03	0,28	26,50
6	www.bcn.cl	80	3	2	5	0,14	0,05	0,27	0,00
7	www.cplt.cl	698	2	2	4	0,13	0,03	0,28	0,00
8	extranet.consejotransparencia.cl	12	2	1	3	0,10	0,03	0,27	0,00
11	productos3.legalpublishing.cl	1	2	0	2	0,00	0,03	0,25	0,00
12	www.redrta.org	2	2	0	2	0,00	0,03	0,25	0,00
18	www.microsoft.com	3	2	0	2	0,00	0,04	0,25	0,00
19	get.adobe.com	2	2	0	2	0,00	0,04	0,25	0,00
20	www.java.com	1	2	0	2	0,00	0,04	0,25	0,00
21	www.openoffice.org	1	2	0	2	0,00	0,04	0,25	0,00
9	200.91.44.244	17	1	2	3	0,13	0,01	0,28	0,00
10	www.interior.gob.cl	28	1	1	2	0,10	0,01	0,27	0,00
13	www.infolobby.cl	25	1	1	2	0,10	0,01	0,27	0,00
14	www.youtube.com	4	1	1	2	0,10	0,01	0,27	0,00
15	labtransparencia.cl	10	1	1	2	0,03	0,01	0,28	0,00
16	www.contraloria.cl	26	1	1	2	0,10	0,01	0,27	0,00
17	goo.gl	2	1	0	1	0,00	0,02	0,25	0,00
22	extranet.cplt.cl	2	1	1	2	0,10	0,02	0,27	0,00
5	www.serviuantofagasta.cl	6	0	1	1	0,03	0,00	0,25	0,00
23	www.bibliotecacongreso.cl	1	0	1	1	0,03	0,00	0,25	0,00
24	en.wikipedia.org	2	0	1	1	0,03	0,00	0,25	0,00
25	www.guiadigital.gob.cl	3	0	1	1	0,10	0,00	0,28	0,00

Tabla 19. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 5

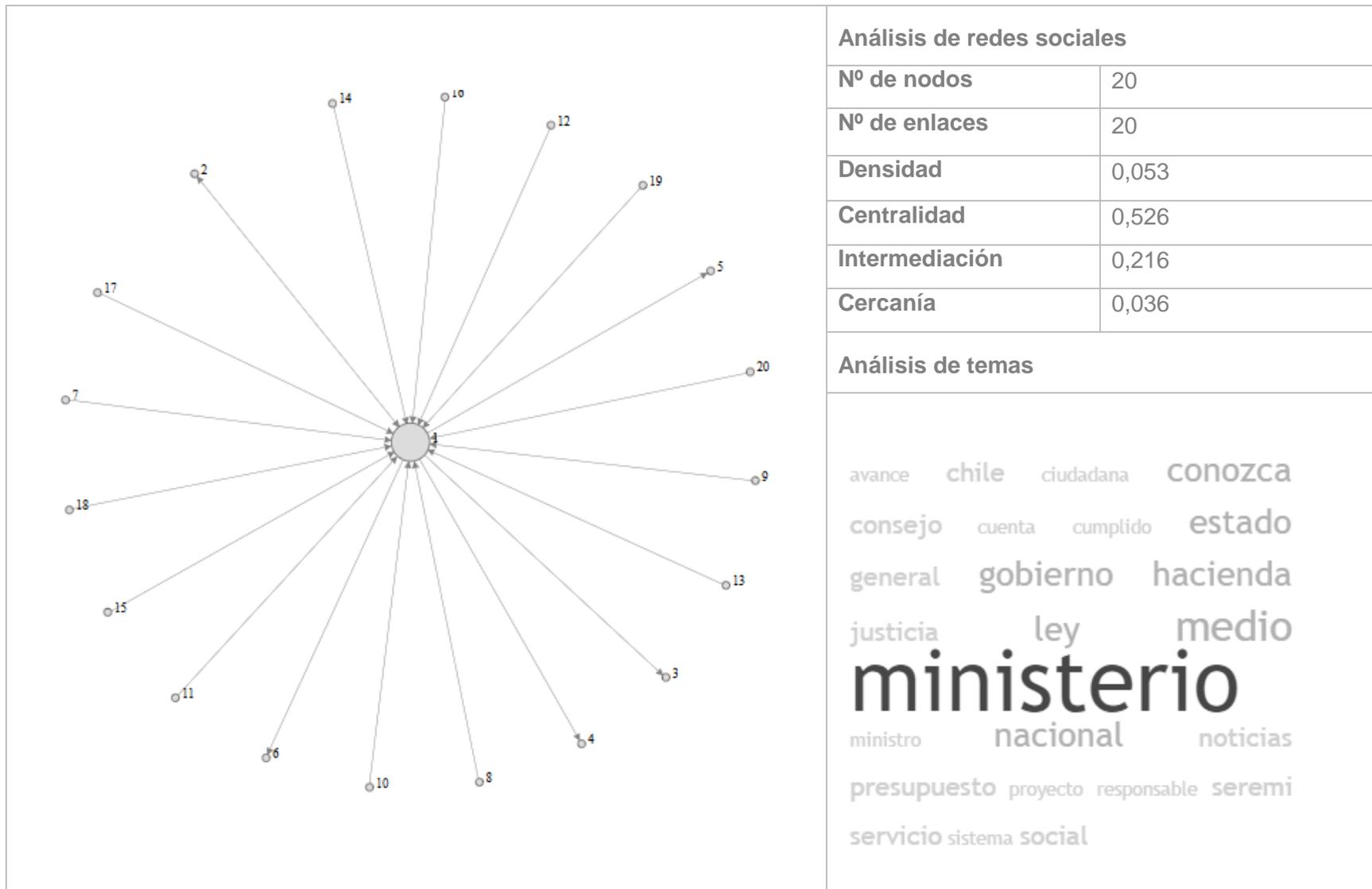


Figura 67. Representación gráfica y análisis de redes sociales y de temas del Clúster 6

Nº	Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
1	www.agendadeprobidad.gob.cl	2	15	5	20	4E-21	1E+00	4E-01	7E+01
2	www.hacienda.gob.cl	2	1	1	2	3E-01	1E-19	4E-01	0E+00
3	www.minjusticia.gob.cl	28	1	0	1	0E+00	1E-19	3E-01	0E+00
4	www.mineduc.gob.cl	25	1	0	1	0E+00	1E-19	3E-01	0E+00
5	www.economia.gob.cl	5	1	0	1	0E+00	1E-19	3E-01	0E+00
6	www.servei.cl	3	1	0	1	0E+00	1E-19	3E-01	0E+00
7	www.soychile.cl	2	0	1	1	3E-01	0E+00	4E-01	0E+00
8	www.hacienda.cl	27	0	1	1	3E-01	0E+00	4E-01	0E+00
9	www.gobernacioniquique.gov.cl	2	0	1	1	3E-01	0E+00	4E-01	0E+00
10	www.gobernacionlimari.gov.cl	1	0	1	1	3E-01	0E+00	4E-01	0E+00
11	www.gobernacioncordillera.gov.cl	2	0	1	1	3E-01	0E+00	4E-01	0E+00
12	www.minhda.cl	4	0	1	1	3E-01	0E+00	4E-01	0E+00
13	www.intendenciaaricaparinacota.gov	1	0	1	1	3E-01	0E+00	4E-01	0E+00
14	www.gobernacionanfeliipe.gov.cl	1	0	1	1	3E-01	0E+00	4E-01	0E+00
15	www.gobernaciontalca.gov.cl	3	0	1	1	3E-01	0E+00	4E-01	0E+00
16	www.gobernacioncolchagua.gov.cl	2	0	1	1	3E-01	0E+00	4E-01	0E+00
17	www.gobernacionquillota.gov.cl	1	0	1	1	3E-01	0E+00	4E-01	0E+00
18	www.gobernacioncapitanprat.gov.cl	1	0	1	1	3E-01	0E+00	4E-01	0E+00
19	www.gobernacionmaipo.gov.cl	1	0	1	1	3E-01	0E+00	4E-01	0E+00
20	www.intendenciaaysen.gov.cl	1	0	1	1	3E-01	0E+00	4E-01	0E+00

Tabla 20. Identificación y análisis de redes sociales de los actores más relevantes del Clúster 6

4.2.2. Análisis de actores

Una vez hemos analizado la red de hiperenlaces en su conjunto y los clústeres que existen en ella procedemos a analizar los actores que forman parte de la red. La metodología que proponíamos para analizar los actores de la red planteaba en primer lugar identificar a los actores, en segundo lugar identificar las redes de vecindario de cada actor y finalmente calcular los niveles de centralidad, intermediación y cercanía de cada uno de ellos.

a) Identificación de actores

VOSON agrupa automáticamente los resultados del análisis que comparten el mismo dominio principal en su URL, pero como indicábamos en la metodología propuesta, en ocasiones, un único actor puede estar representado por varios nodos en la red o dos actores compartir un único nodo. Para evitar esto planteábamos la necesidad de llevar a cabo un procesamiento previo de los resultados que a su vez requería de un análisis cualitativo de las páginas. Este proceso tiene unos costos de tiempo muy altos y consideramos que es posible obtener resultados similares de forma más eficiente utilizando las opciones de visualización de redes de vecindario ofrecidas por VOSON. Uno de los objetivos de esta investigación es aportar una metodología que permita analizar redes temáticas de forma automática o semiautomática por lo que descartamos aplicar un procesamiento cualitativo manual en cualquiera de las fases del análisis.

b) Análisis de redes sociales

Para analizar los nodos desde la perspectiva de las redes sociales proponíamos calcular el grado de centralidad en función de los enlaces entrantes y salientes, el grado de intermediación, el grado de cercanía y el valor de HITS en sus dos versiones *HITS authority* y *HITS hub*. Estas medidas serán calculadas sobre la red absoluta sin depurar, es decir, considerando todos los enlaces entrantes y salientes y no solo los enlaces recíprocos y los coenlaces. Como explicábamos en la propuesta de metodología, tomamos esta decisión porque consideramos que en el estudio de los actores puede ser relevante obtener información sobre sus relaciones no solo con actores afines temáticamente, sino también con actores menos similares o menos relacionados con el

tema de discusión. A continuación, en la Tabla 21, presentamos las medidas de centralidad, intermediación, cercanía y HITS de los diez actores que se pueden considerar más importantes en la red en base a las medidas señaladas.

	Enlaces entrantes	Enlaces salientes	HITS Hub	HITS Authority	Intermediación	Cercanía
Actor 1	320	25	0,079	0,709	185176	0,148
Actor 2	306	7	0,017	0,606	113907	0,143
Actor 3	306	7	0,017	0,606	113907	0,143
Actor 4	201	17	0,042	0,250	104728	0,144
Actor 5	195	23	-	0,046	103510	0,139
Actor 6	133	119	0,134	0,170	199384	0,147
Actor 7	127	0	-	0,075	-	0,125
Actor 8	125	0	-	0,012	-	0,125
Actor 9	106	97	0,012	0,020	173781	0,145
Actor 10	82	25	0,002	0,025	71796,7	0,142

Tabla 21. Medidas de redes sociales de los actores 1-10

A continuación presentamos visualizaciones de las redes de vecindario de los actores que nos permitirán analizar cualitativamente la relación entre las características de cada actor y las medidas de análisis de redes sociales que hemos presentado en la Tabla 21.

c) Visualización de redes de vecindario

Para visualizar las redes de vecindario de los actores hemos empleado la opción de visualización de mapas “*Complete Network*” que ofrece VOSON sobre la red de enlaces entrantes y salientes que hemos identificado previamente. Para identificar las redes de vecindario de cada actor hemos seleccionado los nodos en que estamos interesados para visualizar los enlaces entrantes y salientes de cada uno de ellos. El formato de visualización de VOSON nos permite obtener información de los nombres de los nodos pero no es posible insertar ese tipo de gráficos en un documento de estas características. En el caso de la representación de las redes de vecindario de los actores, la numeración de los nodos que habíamos empleado en los clústers hacía imposible la correcta visualización de las redes por lo que hemos optado por adjuntar una tabla con los actores

más relevantes que crean o reciben un enlace hacia o desde el actor que se analiza en cada caso. Se ha añadido además las informaciones que consideramos más importantes sobre cada actor: nombre, número de páginas que contiene su sitio, número de enlaces entrantes, número de enlaces salientes, número de enlaces totales, nivel HITS Hub, nivel HITS Authority, nivel de cercanía y nivel de intermediación.

Este tipo de representación nos permite visualizar en detalle los enlaces entrantes, salientes y recíprocos de cada actor y así poder inferir el tipo de relaciones directas que presenta con el resto de nodos de la red. Las medidas relacionadas con el análisis de redes sociales nos ofrecen algunas particularidades de cada actor y de su papel en el funcionamiento de la red. Los actores analizados han sido seleccionados preferenciando su nivel de centralidad basado en el número de enlaces entrantes. Sin embargo, también podemos extraer información relevante en las medidas de prestigio que nos ofrece el cálculo de HITS Hub y HITS Authority, en las de intermediación y en las de cercanía.

En la Tabla 21, vemos por ejemplo que el actor 6 es el que presenta un mayor nivel de intermediación a pesar de no ser uno de los actores que más enlaces entrantes recibe o que mayor nivel de prestigio basado en HITS Authority presenta. En el caso de la cercanía vemos por ejemplo que el actor 9 presenta un nivel de cercanía con el resto de nodos de la red que solo es superado por el actor 1 que triplica el número de enlaces entrantes y multiplica casi por 30 el valor HITS Authority.

El objetivo de esta investigación no es analizar en detalle todos los actores de la red temática de la transparencia en Chile, sino mostrar que la metodología propuesta nos permite alcanzar los objetivos planteados para el análisis de redes temáticas en la web y concretamente para el análisis de los actores que forman parte de ella. Consideramos que los resultados obtenidos al analizar los diez actores seleccionados nos permiten afirmar que la metodología propuesta es adecuada para alcanzar los objetivos de identificación, relación con el resto de la red y descripción de los marcos de referencia de un actor.

A continuación, en las Figuras 63, 64, 65, 66, 67, 68, 69, 70, 71 y 72 y las tablas correspondientes a cada una de ellas, vemos la representación de las redes de vecindario de los principales actores de la red y su relación con los actores más relevantes.

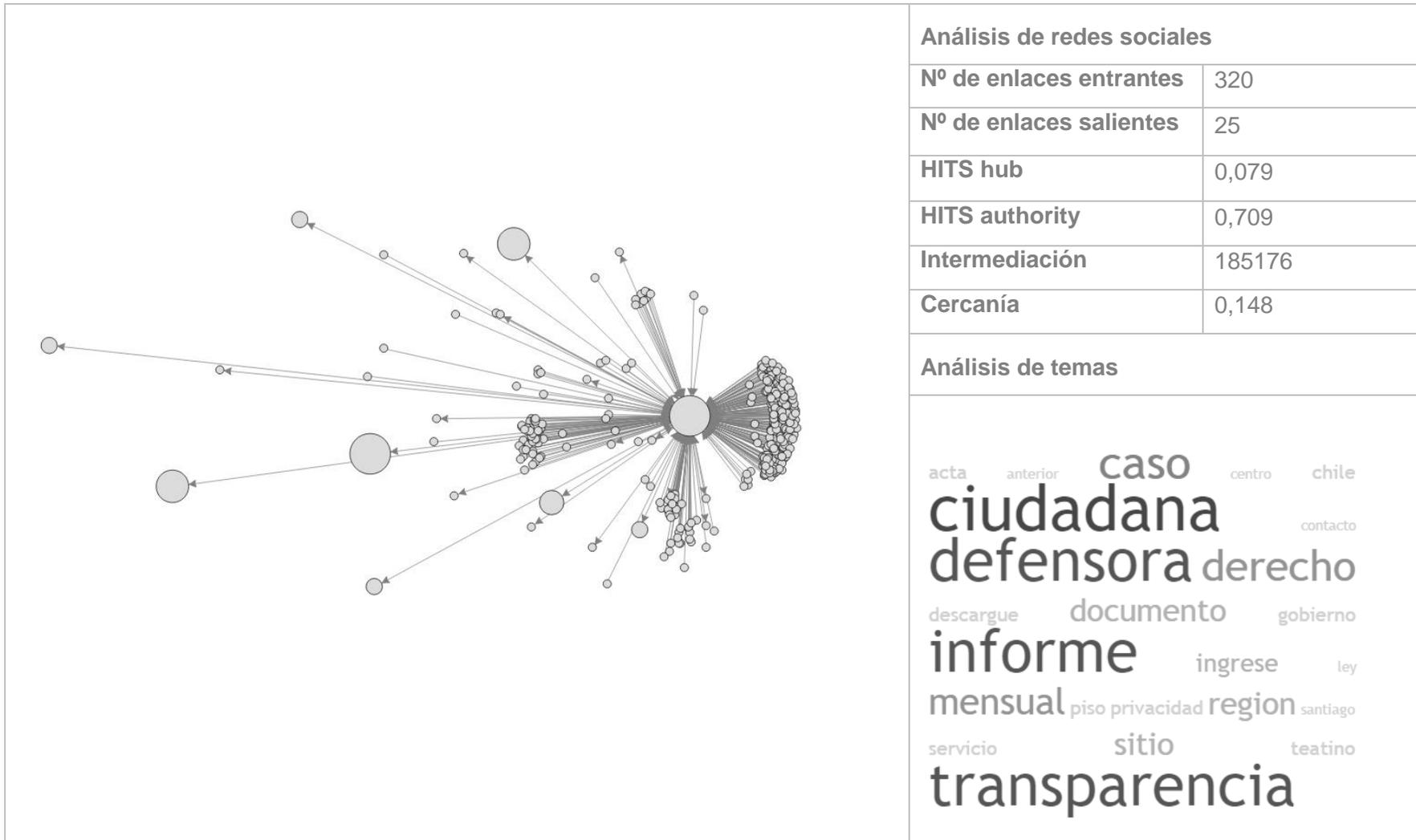


Figura 68. Representación gráfica de la red de vecindario, análisis de redes sociales y análisis de temas del actor Comisión Defensora Ciudadana y Transparencia

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.educatransparencia.cl	134	139	7	146	0,03	0,53	0,17	13065,10
www.consejotransparencia.cl	432	108	16	124	0,06	0,38	0,17	15307,40
www.gobiernotransparentechile.cl	1674	97	80	177	0,19	0,33	0,18	37641,80
www.ciudadano inteligente.org	47	65	11	76	0,00	0,07	0,16	10953,10
www.datos.gob.cl	80	56	11	67	0,10	0,21	0,18	3186,17
www.chiletransparente.cl	227	39	7	46	0,00	0,06	0,17	6324,37
www.leychile.cl	1093	37	0	37	0,00	0,12	0,13	0,00
www.probono.cl	409	37	37	74	0,04	0,04	0,17	20227,50
www.portaltransparencia.cl	493	30	20	50	0,11	0,10	0,18	20146,00
www.proacceso.cl	22	26	2	28	0,00	0,04	0,16	2381,42
www.agendadeprobidad.gob.cl	2	20	8	28	0,01	0,07	0,17	1756,99
www.leydetransparencia.cl	62	13	14	27	0,09	0,05	0,18	2064,37
www.chileatiende.cl	47	7	1	8	0,04	0,05	0,17	273,69
www.minsegres.gob.cl	9	5	5	10	0,08	0,03	0,18	743,94
transparencia.minsegres.gob.cl	6	3	1	4	0,03	0,03	0,16	116,03
dpi.minsegres.gob.cl	3	3	0	3	0,00	0,02	0,13	0,00
www.gob.cl	44	2	2	4	0,04	0,02	0,17	150,65
participacionciudadana.minsegres.gob.cl	32	2	3	5	0,07	0,01	0,17	0,00
www.opengovpartnership.org	5	2	1	3	0,00	0,01	0,15	522,05
www.bibliotecacongreso.cl	1	1	1	2	0,01	0,01	0,13	0,00

Tabla 22. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Comisión Defensora Ciudadana y Transparencia

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.gobiernotransparentechile.cl	1674	97	80	177	0,19	0,33	0,18	37641,80
www.datos.gob.cl	80	56	11	67	0,10	0,21	0,18	3186,17
www.leydetransparencia.cl	62	13	14	27	0,09	0,05	0,18	2064,37
www.dt.gob.cl	9265	7	12	19	0,09	0,03	0,17	1395,31
www.chileatiende.cl	47	7	1	8	0,04	0,05	0,17	273,69
www.subtrans.gob.cl	25	5	2	7	0,07	0,02	0,17	131,21
www.minsegpres.gob.cl	9	5	5	10	0,08	0,03	0,18	743,94
www.comisiondefensoraciudadana.cl	1	3	1	4	0,04	0,01	0,17	584,60
www.hacienda.gob.cl	2	2	2	4	0,04	0,01	0,17	203,31
web.minsal.cl	26	2	1	3	0,04	0,02	0,17	105,64
www.gob.cl	44	2	2	4	0,04	0,02	0,17	150,65
participacionciudadana.minsegpres.gob.cl	32	2	3	5	0,07	0,01	0,17	0,00
www.modernizacion.gob.cl	25	2	3	5	0,07	0,01	0,17	0,00
www.ssffaa.gob.cl	2	1	1	2	0,04	0,01	0,17	0,00
www.ssffaa.cl	27	1	1	2	0,04	0,01	0,17	0,00
www.ind.cl	25	1	1	2	0,04	0,01	0,17	0,00
www.cultura.gob.cl	26	1	2	3	0,06	0,01	0,17	0,00
www.sec.cl	9	1	2	3	0,07	0,01	0,17	0,00
www.dsp.gob.cl	1	1	1	2	0,04	0,01	0,17	0,00
www.subtel.gob.cl	6	1	1	2	0,04	0,01	0,17	0,00
www.direcon.gob.cl	7	1	3	4	0,10	0,01	0,17	0,00
www.minrel.gov.cl	18	1	2	3	0,07	0,01	0,17	0,00
www.minrel.gob.cl	27	1	2	3	0,07	0,01	0,17	0,00
www.carabineros.cl	9	1	2	3	0,07	0,01	0,17	0,00
www.cchen.cl	20	1	3	4	0,07	0,01	0,17	161,67

Tabla 23. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Comisión Defensora Ciudadana y Transparencia

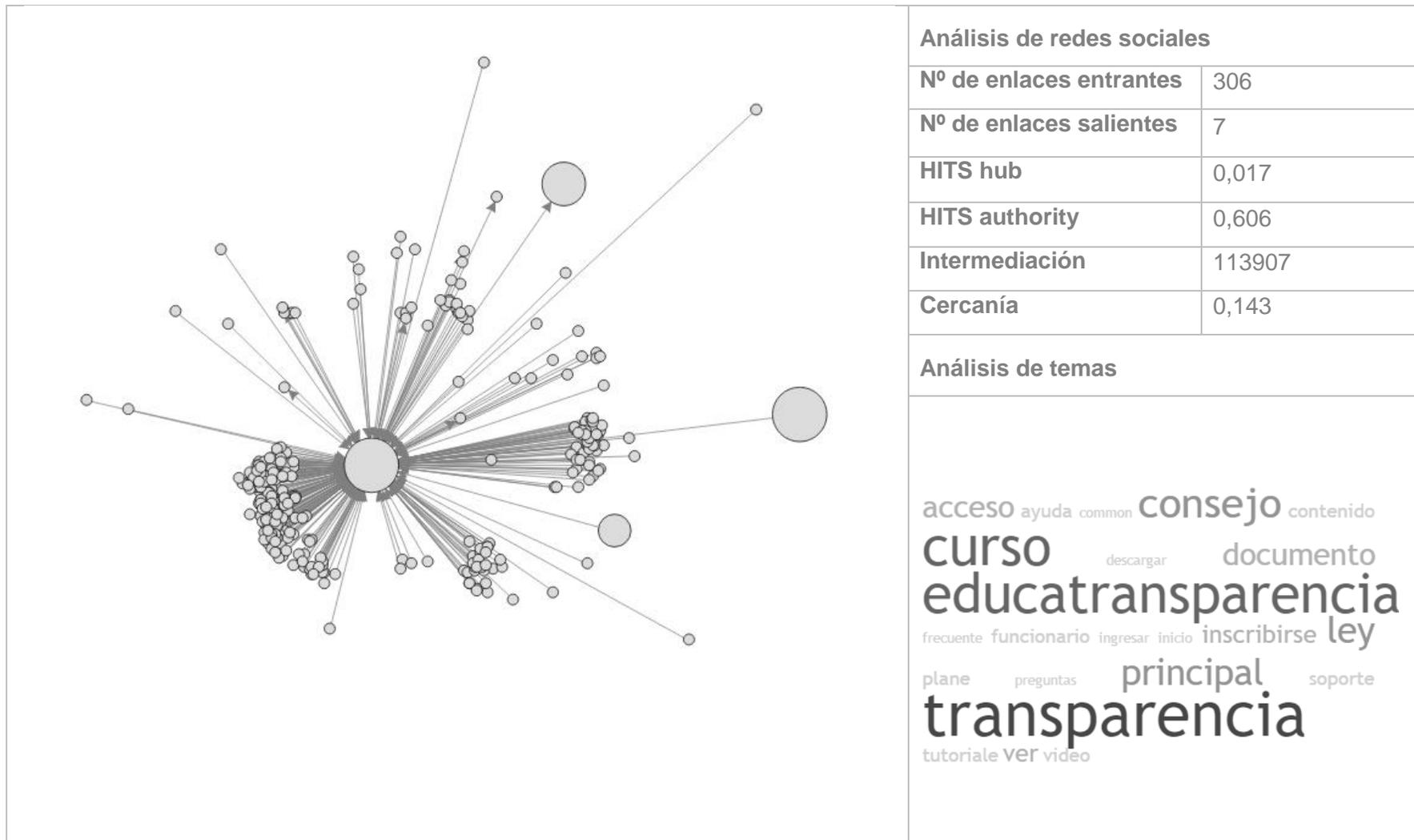


Figura 69. Representación gráfica de la red de vecindario, análisis de redes sociales y análisis de temas del actor Educatransparencia

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.consejotransparencia.cl	432	108	16	124	0,06	0,38	0,17	15307,40
www.portaltransparencia.cl	493	30	20	50	0,11	0,10	0,18	20146,00
www.cplt.cl	698	3	2	5	0,03	0,01	0,17	0,00

Tabla 24. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Educatransparencia

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.cdc.gob.cl	57	136	22	158	0,14	0,58	0,18	28946,30
www.consejotransparencia.cl	432	108	16	124	0,06	0,38	0,17	15307,40
www.gobiernotransparentechile.cl	1674	97	80	177	0,19	0,33	0,18	37641,80
www.portaltransparencia.cl	493	30	20	50	0,11	0,10	0,18	20146,00
www.mineduc.cl	379	18	16	34	0,06	0,05	0,17	4341,79
transparencia.dga.cl	301	6	22	28	0,09	0,03	0,17	4075,22
www.subtrans.gob.cl	25	5	2	7	0,07	0,02	0,17	131,21
www.dipres.gob.cl	65	5	1	6	0,03	0,03	0,16	0,53
www.gobiernotransparente.cl	35	5	1	6	0,03	0,03	0,16	0,53
www.uchile.cl	7	3	5	8	0,04	0,01	0,16	2361,33
transparencia.minsepres.gob.cl	6	3	1	4	0,03	0,03	0,16	116,03
www.registrocivil.cl	10	2	1	3	0,03	0,01	0,16	114,52
transparencia.dcyf.cl	6	2	1	3	0,03	0,02	0,16	0,00
transparencia.mop.cl	10	2	1	3	0,03	0,02	0,16	0,00
transparencia.dgop.cl	4	2	1	3	0,03	0,02	0,16	0,00
extranet.consejotransparencia.cl	12	2	2	4	0,06	0,01	0,16	0,00
www.interior.gob.cl	28	2	2	4	0,06	0,00	0,16	118,10
www.contraloria.cl	26	2	3	5	0,06	0,01	0,16	275,52
transparenciaactiva.cultura.gob.cl	8	1	1	2	0,03	0,01	0,16	0,00

Tabla 25. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Educatransparencia

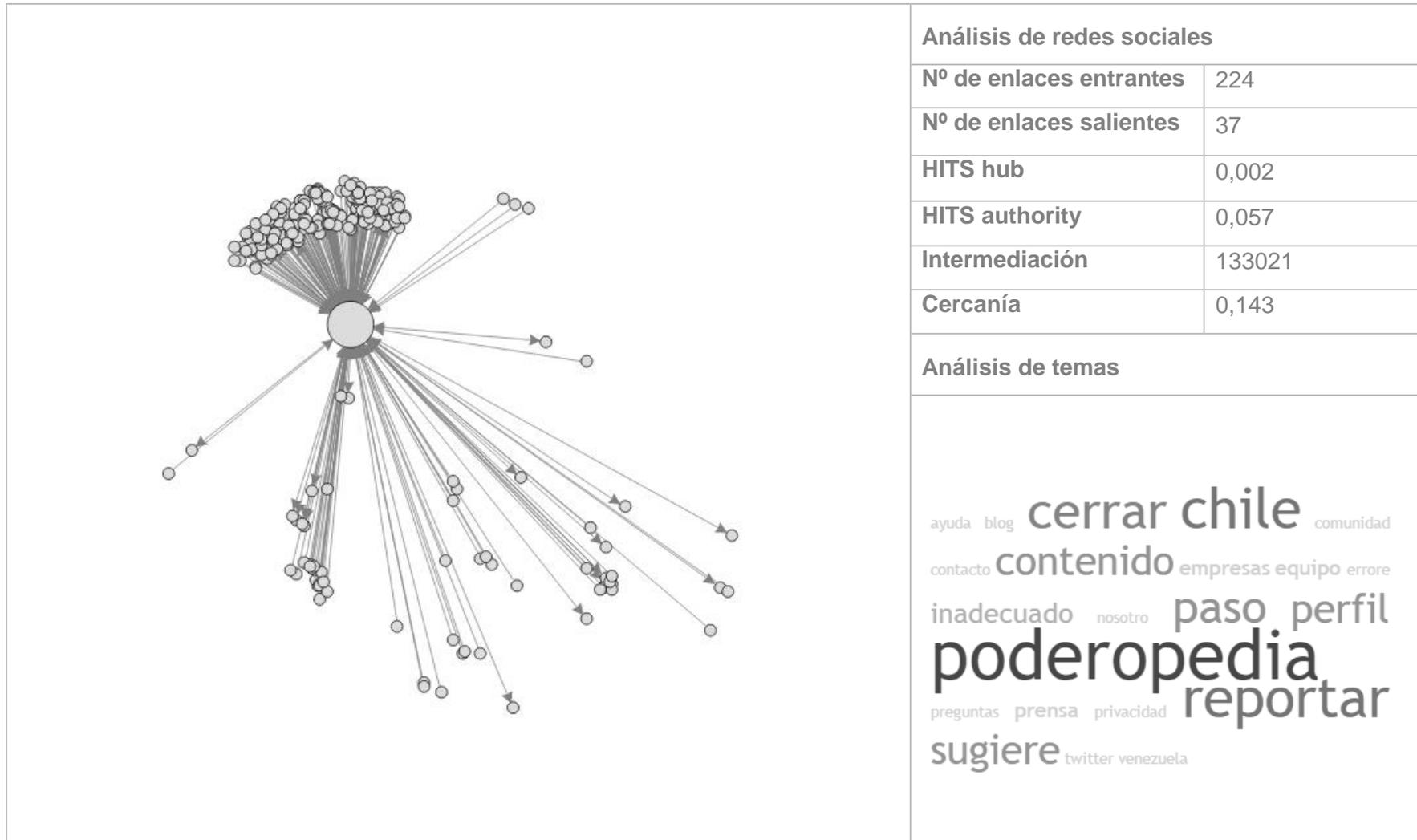


Figura 70. Representación gráfica de la red de vecindario, análisis de redes sociales y análisis de temas del actor Poderopedia

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.mercadopublico.cl	59	7	0	7	0,00	0,03	0,13	0,00
www.registrocivil.cl	10	2	1	3	0,03	0,01	0,16	114,52
www.minjusticia.gob.cl	28	3	1	4	0,02	0,01	0,16	189,18
www.cooperativa.cl	5	1	1	2	0,02	0,00	0,17	1712,39
www.theclinic.cl	8	1	3	4	0,01	0,00	0,16	476,07
radio.uchile.cl	5	1	1	2	0,00	0,00	0,15	343,47
www.latercera.com	11	1	3	4	0,01	0,00	0,16	387,13
www.lasegunda.com	4	1	1	2	0,00	0,00	0,16	0,00
www.capital.cl	2	1	1	2	0,02	0,00	0,17	1712,39
www.elmostrador.cl	14	1	4	5	0,07	0,00	0,17	1986,39
www.24horas.cl	2	1	1	2	0,00	0,00	0,15	343,47
www.lanacion.cl	4	1	2	3	0,03	0,00	0,16	179,44
www.transparencia.org.ve	2	1	1	2	0,00	0,00	0,15	447,36
www.startupchile.org	6	1	1	2	0,00	0,00	0,16	0,00
www.knightfoundation.org	24	1	2	3	0,01	0,00	0,16	343,47
www.icfj.org	51	1	2	3	0,01	0,00	0,16	343,47
www.niemanlab.org	32	1	2	3	0,01	0,00	0,16	343,47
thenextweb.com	3	1	1	2	0,00	0,00	0,16	0,00
blog.poderopedia.org	56	1	2	3	0,01	0,00	0,16	343,47

Tabla 26. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Poderopedia

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.inap.uchile.cl	1	2	2	3	0,01	0,00	0,16	1309,85
www.lasegunda.com	1	1	4	2	0,00	0,00	0,16	0,00
www.elmostrador.cl	1	4	14	5	0,07	0,00	0,17	1986,39
www.lanacion.cl	1	2	4	3	0,03	0,00	0,16	179,44
www.startupchile.org	1	1	6	2	0,00	0,00	0,16	0,00
www.knightfoundation.org	1	2	24	3	0,01	0,00	0,16	343,47
www.icfj.org	1	2	51	3	0,01	0,00	0,16	343,47
www.niemanlab.org	1	2	32	3	0,01	0,00	0,16	343,47
thenextweb.com	1	1	3	2	0,00	0,00	0,16	0,00
blog.poderopedia.org	1	2	56	3	0,01	0,00	0,16	343,47
www.biobiochile.cl	0	3	5	3	0,06	0,00	0,18	0,00
www.lanacion.com.ar	0	2	3	2	0,03	0,00	0,17	0,00
www.elquintopoder.cl	0	6	14	6	0,09	0,00	0,18	0,00
www.conexionsocial.cl	0	2	11	2	0,03	0,00	0,16	0,00
blogs.lanacion.com.ar	0	3	26	3	0,03	0,00	0,17	0,00
www.elciudadano.cl	0	6	17	6	0,05	0,00	0,17	0,00
www.puroperiodismo.cl	0	3	3	3	0,03	0,00	0,17	0,00

Tabla 27. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Poderopedia

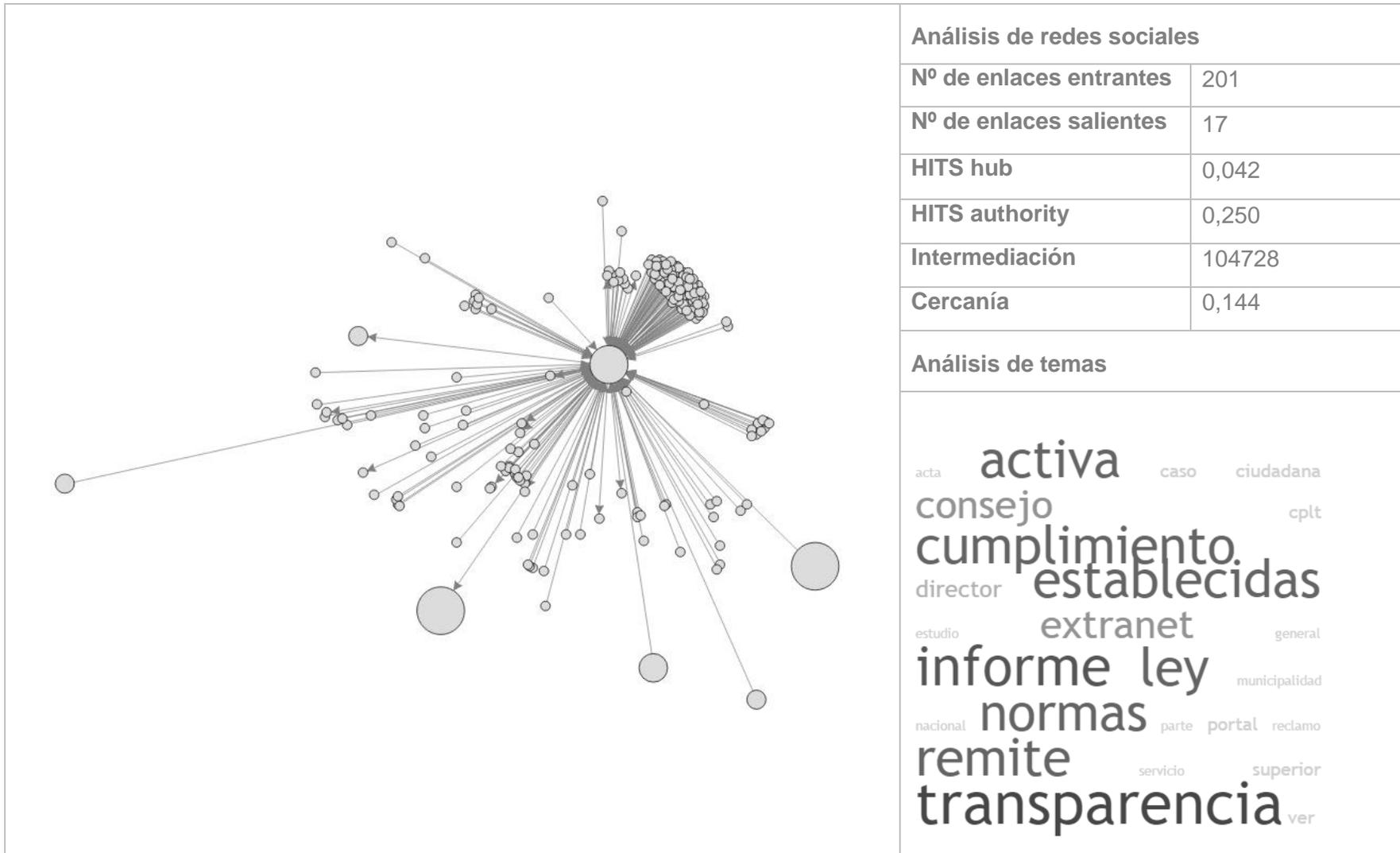


Figura 71. Representación gráfica de la red de vecindario, análisis de redes sociales y análisis de temas del actor Consejo de Transparencia

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.educatransparencia.cl	134	139	7	146	0,03	0,53	0,17	13065,10
www.leychile.cl	1093	37	0	37	0,00	0,12	0,13	0,00
www.portaltransparencia.cl	493	30	20	50	0,11	0,10	0,18	20146,00
www.bcn.cl	80	9	7	16	0,06	0,05	0,18	13828,30
www.gobiernotransparente.cl	35	5	1	6	0,03	0,03	0,16	0,53
www.cplt.cl	698	3	2	5	0,03	0,01	0,17	0,00
extranet.consejotransparencia.cl	12	2	2	4	0,06	0,01	0,16	0,00
www.interior.gob.cl	28	2	2	4	0,06	0,00	0,16	118103,00
productos3.legalpublishing.cl	1	2	0	2	0,00	0,01	0,13	0,00
www.redrta.org	2	2	0	2	0,00	0,01	0,13	0,00
www.contraloria.cl	26	2	3	5	0,06	0,01	0,16	275523,00
www.infolobby.cl	25	1	1	2	0,02	0,00	0,16	0,00
labtransparencia.cl	10	1	1	2	0,01	0,00	0,16	0,00

Tabla 28. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Consejo de Transparencia

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.educatransparencia.cl	139	7	134	146	0,03	0,53	0,17	13065,10
www.cdc.gob.cl	136	22	57	158	0,14	0,58	0,18	28946,30
www.gobiernotransparentechile.cl	97	80	1674	177	0,19	0,33	0,18	37641,80
www.datos.gob.cl	56	11	80	67	0,10	0,21	0,18	3186,17
www.probono.cl	37	37	409	74	0,04	0,04	0,17	20227,50
www.portaltransparencia.cl	30	20	493	50	0,11	0,10	0,18	20146,00
www.bcn.cl	9	7	80	16	0,06	0,05	0,18	13828,30
www.dt.gob.cl	7	12	9265	19	0,09	0,03	0,17	1395,31
transparencia.dga.cl	6	22	301	28	0,09	0,03	0,17	4075,22
www.transparenciamaipu.cl	4	9	184	13	0,05	0,01	0,17	256,18
www.minjusticia.gob.cl	3	1	28	4	0,02	0,01	0,16	189,18
www.cplt.cl	3	2	698	5	0,03	0,01	0,17	0,00
extranet.consejotransparencia.cl	2	2	12	4	0,06	0,01	0,16	0,00
www.interior.gob.cl	2	2	28	4	0,06	0,00	0,16	118103,00
www.contraloria.cl	2	3	26	5	0,06	0,01	0,16	275523,00
www.cultura.gob.cl	1	2	26	3	0,06	0,01	0,17	0,00
www.direcon.gob.cl	1	3	7	4	0,10	0,01	0,17	0,00
www.subdere.gov.cl	1	2	11	3	0,06	0,01	0,16	0,00
transparencia.cntv.cl	1	2	7	3	0,06	0,01	0,16	0,00
www.infolobby.cl	1	1	25	2	0,02	0,00	0,16	0,00
www.gobiernoabierto.gob.cl	1	4	82	5	0,10	0,01	0,18	225,36
extranet.cplt.cl	1	1	2	2	0,02	0,01	0,16	0,00
www.elmostrador.cl	1	4	14	5	0,07	0,00	0,17	1986,39
www.lanacion.cl	1	2	4	3	0,03	0,00	0,16	179,44
consorcioporlatransparencia.cl	1	5	26	6	0,04	0,00	0,17	1695,42
www.cajmetro.cl	1	2	27	3	0,06	0,00	0,17	584,60

Tabla 29. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Consejo Transparencia

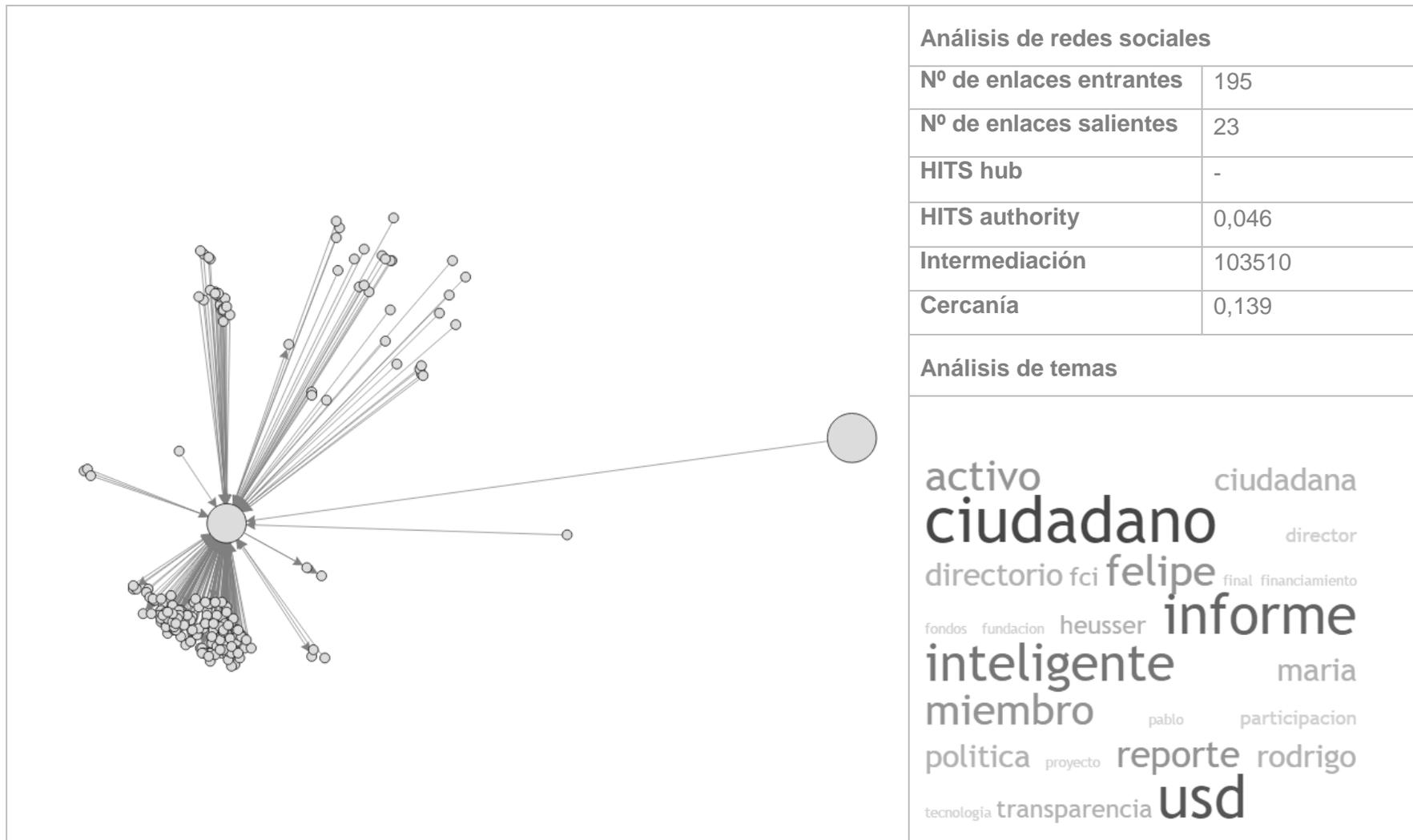


Figura 72. Representación gráfica de la red de vecindario y análisis de redes sociales de Ciudadano Inteligente

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
blog.ciudadanointeligente.org	7	1	1	2	0,00	0,00	0,15	0,00
poplus.org	13	1	1	2	0,00	0,00	0,15	0,00
hayacuerdo.ciudadanointeligente.org	2	1	1	2	0,00	0,00	0,15	0,00
www.deldichoalhecho.cl	5	1	1	2	0,00	0,00	0,15	0,00
www.transparencialegislativa.org	3	1	2	3	0,01	0,00	0,16	417,02
votainteligente.cl	10	1	1	2	0,00	0,00	0,15	0,00
quientefinancia.cl	3	1	2	3	0,01	0,00	0,16	575,77
www.candideit.org	25	1	1	2	0,00	0,00	0,15	0,00
congresoabierto.cl	25	1	1	2	0,00	0,00	0,15	0,00
leydelobby.cl	6	1	2	3	0,01	0,00	0,16	575,77

Tabla 30. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Ciudadano Inteligente

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.cdc.gob.cl	57	136	22	158	0,14	0,58	0,18	28946,30
www.opengovpartnership.org	5	2	1	3	0,00	0,01	0,15	522052,00
www.inap.uchile.cl	2	1	2	3	0,01	0,00	0,16	1309,85
blog.ciudadanointeligente.org	7	1	1	2	0,00	0,00	0,15	0,00
poplus.org	13	1	1	2	0,00	0,00	0,15	0,00
hayacuerdo.ciudadanointeligente.org	2	1	1	2	0,00	0,00	0,15	0,00
www.deldichoalhecho.cl	5	1	1	2	0,00	0,00	0,15	0,00
www.transparencialegislativa.org	3	1	2	3	0,01	0,00	0,16	417,02
votainteligente.cl	10	1	1	2	0,00	0,00	0,15	0,00
www.candideit.org	25	1	1	2	0,00	0,00	0,15	0,00
congresoabierto.cl	25	1	1	2	0,00	0,00	0,15	0,00
www.theclinic.cl	8	1	3	4	0,01	0,00	0,16	476,07
radio.uchile.cl	5	1	1	2	0,00	0,00	0,15	343,47
www.latercera.com	11	1	3	4	0,01	0,00	0,16	387,13
www.24horas.cl	2	1	1	2	0,00	0,00	0,15	343,47
www.knightfoundation.org	24	1	2	3	0,01	0,00	0,16	343,47
www.icfj.org	51	1	2	3	0,01	0,00	0,16	343,47
www.niemanlab.org	32	1	2	3	0,01	0,00	0,16	343,47
blog.poderopedia.org	56	1	2	3	0,01	0,00	0,16	343,47
consorcioporlatransparencia.cl	26	1	5	6	0,04	0,00	0,17	1695,42
www.ayudar.cl	26	1	1	2	0,00	0,00	0,15	183,60
www.elquintopoder.cl	14	0	6	6	0,09	0,00	0,18	0,00
www.elmorrocotudo.cl	8	0	2	2	0,03	0,00	0,16	0,00
blogs.lanacion.com.ar	26	0	3	3	0,03	0	0,17	0,00
tenred.cl	9	0	3	3	0,03	0	0,16	0,00
www.alejandrobarrros.com	5	0	2	2	0,03	0	0,16	0,00

Tabla 31. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Ciudadano Inteligente

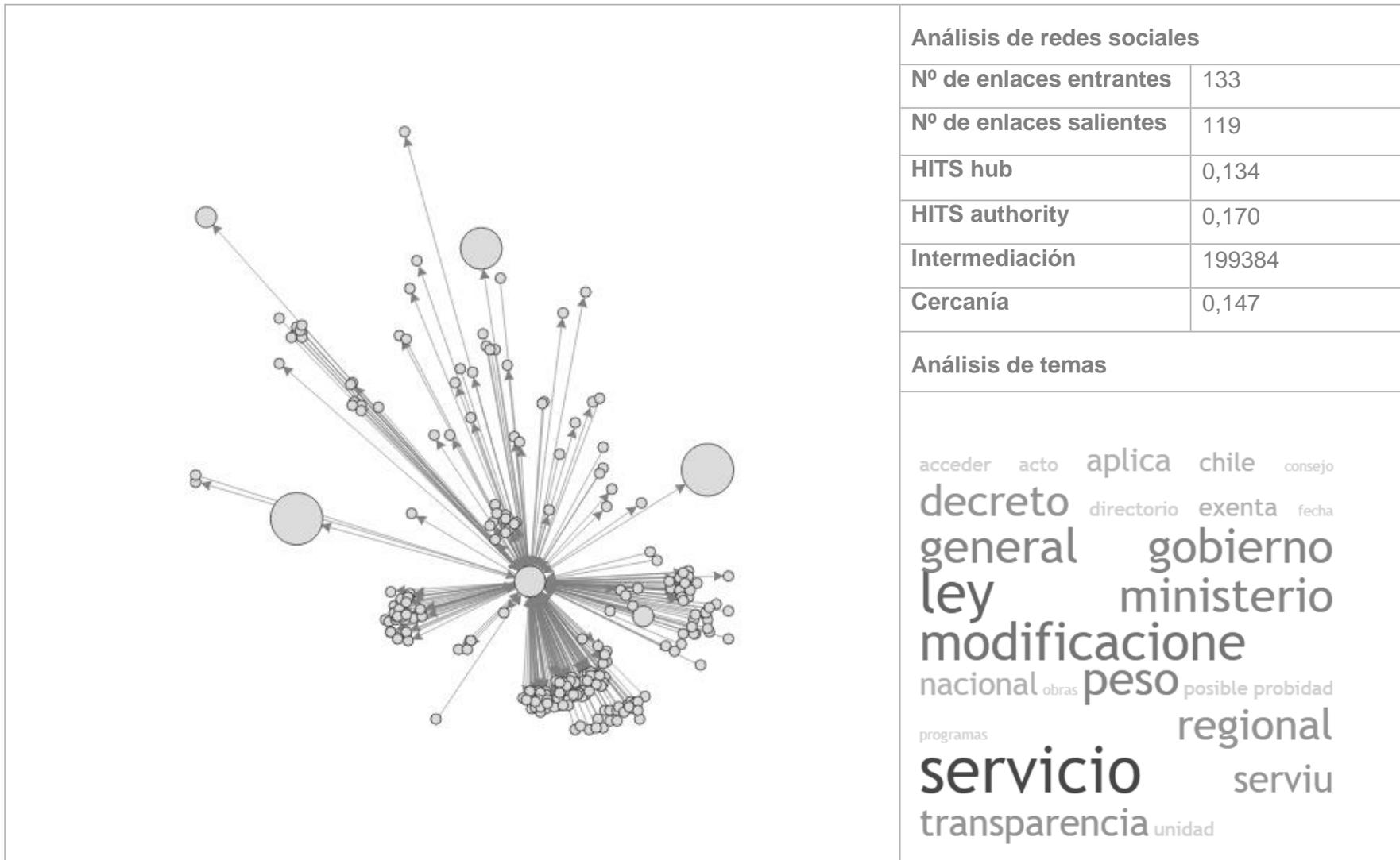


Figura 73. Representación gráfica de la red de vecindario y análisis de redes sociales de Gobierno Transparente

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.educatransparencia.cl	134	139	7	146	0,03	0,53	0,17	13065,10
www.cdc.gob.cl	57	136	22	158	0,14	0,58	0,18	28946,30
www.consejotransparencia.cl	432	108	16	124	0,06	0,38	0,17	15307,40
www.leychile.cl	1093	37	0	37	0,00	0,12	0,13	0,00
www.portaltransparencia.cl	493	30	20	50	0,11	0,10	0,18	20146,00
www.mineduc.cl	379	18	16	34	0,06	0,05	0,17	4341,79
www.bcn.cl	80	9	7	16	0,06	0,05	0,18	13828,30
www.dt.gob.cl	9265	7	12	19	0,09	0,03	0,17	1395,31
www.chileatiende.cl	47	7	1	8	0,04	0,05	0,17	273689,00
www.mercadopublico.cl	59	7	0	7	0,00	0,03	0,13	0,00
transparencia.dga.cl	301	6	22	28	0,09	0,03	0,17	4075,22
www.subtrans.gob.cl	25	5	2	7	0,07	0,02	0,17	131,21
www.dipres.gob.cl	65	5	1	6	0,03	0,03	0,16	0,53
www.registros19862.cl	35	5	0	5	0,00	0,03	0,13	0,00
www.gobiernotransparente.gob.cl	3	5	0	5	0,00	0,03	0,13	0,00
www.minsegres.gob.cl	9	5	5	10	0,08	0,03	0,18	743,94
www.gobiernotransparente.cl	35	5	1	6	0,03	0,03	0,16	0,53
www.gobiernotransparente.gov.cl	3	4	0	4	0,00	0,03	0,13	0,00
transparencia.minsegres.gob.cl	6	3	1	4	0,03	0,03	0,16	116,03
www.minjusticia.gob.cl	28	3	1	4	0,02	0,01	0,16	189,18
www.registrocivil.cl	10	2	1	3	0,03	0,01	0,16	114,52
transparencia.dcyf.cl	6	2	1	3	0,03	0,02	0,16	0,00
www.dcyf.cl	2	2	0	2	0,00	0,02	0,13	0,00
www.hacienda.gob.cl	2	2	2	4	0,04	0,01	0,17	203,31
www.sernam.cl	7	2	0	2	0,00	0,01	0,13	0,00

Tabla 32. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Gobierno Transparente

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.cdc.gob.cl	136	22	57	158	0,14	0,58	0,18	28946,30
www.leydetransparencia.cl	13	14	62	27	0,09	0,05	0,18	2064,37
www.portaltransparencia.cl	30	20	493	50	0,11	0,10	0,18	20146,00
www.bcn.cl	9	7	80	16	0,06	0,05	0,18	13828,30
www.datos.gob.cl	56	11	80	67	0,10	0,21	0,18	3186,17
www.minsegpres.gob.cl	5	5	9	10	0,08	0,03	0,18	743,94
www.cchen.cl	1	3	20	4	0,07	0,01	0,17	161,67
portal.mma.gob.cl	1	2	3	3	0,03	0,01	0,17	0,00
participacionciudadana.minsegpres.gob.cl	2	3	32	5	0,07	0,01	0,17	0,00
www.participacionciudadana.gob.cl	1	1	2	2	0,02	0,01	0,17	0,00
www.modernizacion.gob.cl	2	3	25	5	0,07	0,01	0,17	0,00
www.gobiernoabierto.gob.cl	1	4	82	5	0,10	0,01	0,18	225,36
apps.gob.cl	1	3	25	4	0,07	0,01	0,17	0,00
www.ogp.cl	1	4	3	5	0,09	0,01	0,18	227,67
www.cooperativa.cl	1	1	5	2	0,02	0,00	0,17	1712,39
www.capital.cl	1	1	2	2	0,02	0,00	0,17	1712,39
www.pac.dsp.gob.cl	0	3	25	3	0,07	0,00	0,17	0,00
participacionciudadana.sag.gob.cl	0	3	12	3	0,07	0,00	0,17	0,00
www.observatoriodigital.gob.cl	0	3	25	3	0,07	0,00	0,17	0,00
participacionciudadana.mintrab.gob.cl	0	3	20	3	0,07	0,00	0,17	0,00
participacionciudadana.minvu.cl	0	3	25	3	0,07	0,00	0,17	0,00
pac.consejodirectores.cl	0	2	17	2	0,03	0,00	0,17	0,00
participacionciudadana.inapi.cl	0	3	25	3	0,07	0,00	0,17	0,00
idealisreports.cl	0	2	4	2	0,02	0,00	0,18	0,00
pac.superir.gob.cl	0	2	21	2	0,03	0,00	0,17	0,00
pac.conadi.cl	0	2	15	2	0,03	0,00	0,17	0,00

Tabla 33. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente

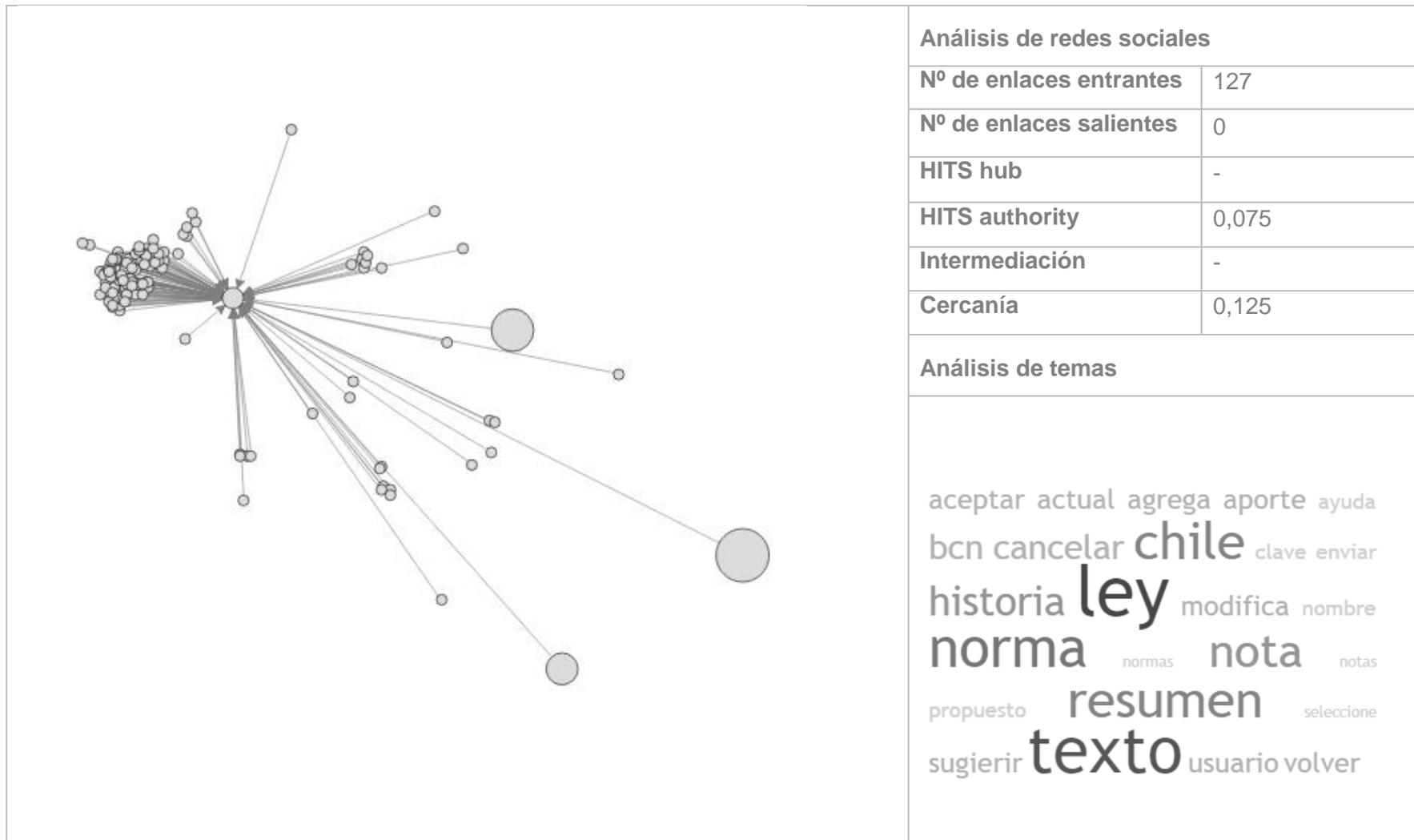


Figura 74. Representación gráfica de la red de vecindario y análisis de redes sociales de Ley Chile

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.cdc.gob.cl	57	136	22	158	0,14	0,58	0,18	28946,30
www.consejotransparencia.cl	432	108	16	124	0,06	0,38	0,17	15307,40
www.gobiernotransparentechile.cl	1674	97	80	177	0,19	0,33	0,18	37641,80
www.portaltransparencia.cl	493	30	20	50	0,11	0,10	0,18	20146,00
www.mineduc.cl	379	18	16	34	0,06	0,05	0,17	4341,79
www.bcn.cl	80	9	7	16	0,06	0,05	0,18	13828,30
www.dt.gob.cl	9265	7	12	19	0,09	0,03	0,17	1395,31
transparencia.dga.cl	301	6	22	28	0,09	0,03	0,17	4075,22
www.transparenciamaipu.cl	184	4	9	13	0,05	0,01	0,17	256,18
www.serviufogasta.cl	6	1	1	2	0,01	0,01	0,13	0,00
www.bibliotecacongreso.cl	1	1	1	2	0,01	0,01	0,13	0,00
en.wikipedia.org	2	1	1	2	0,01	0,00	0,13	36,63
www.theglobe.org	3	0	2	2	0,03	0,00	0,17	0,00
www.the-globe.com	4	0	2	2	0,03	0,00	0,17	0,00
collabmarket.org	17	0	2	2	0,03	0,00	0,17	0,00
www.loncochetransparente.cl	23	0	2	2	0,03	0,00	0,16	0,00
www.elciudadano.cl	17	0	6	6	0,05	0,00	0,17	0,00
www.muniercilla.cl	14	0	2	2	0,03	0,00	0,16	0,00
www.municipalidadcolbun.cl	29	0	2	2	0,03	0,00	0,16	0,00
www.constitucion.cl	5	0	4	4	0,07	0,00	0,17	0,00
www.machali.cl	7	0	2	2	0,03	0,00	0,16	0,00
www.icdt.cl	10	0	3	3	0,05	0,00	0,17	0,00
topsites.cl	16	0	5	5	0,07	0,00	0,17	0,00
www.padrelascasas.cl	1	0	2	2	0,03	0,00	0,16	0,00
www.serviu7.cl	2	0	2	2	0,03	0,00	0,16	0,00
www.municipiotransparentes.cl	47	0	3	3	0,05	0,00	0,17	0,00

Tabla 34. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente

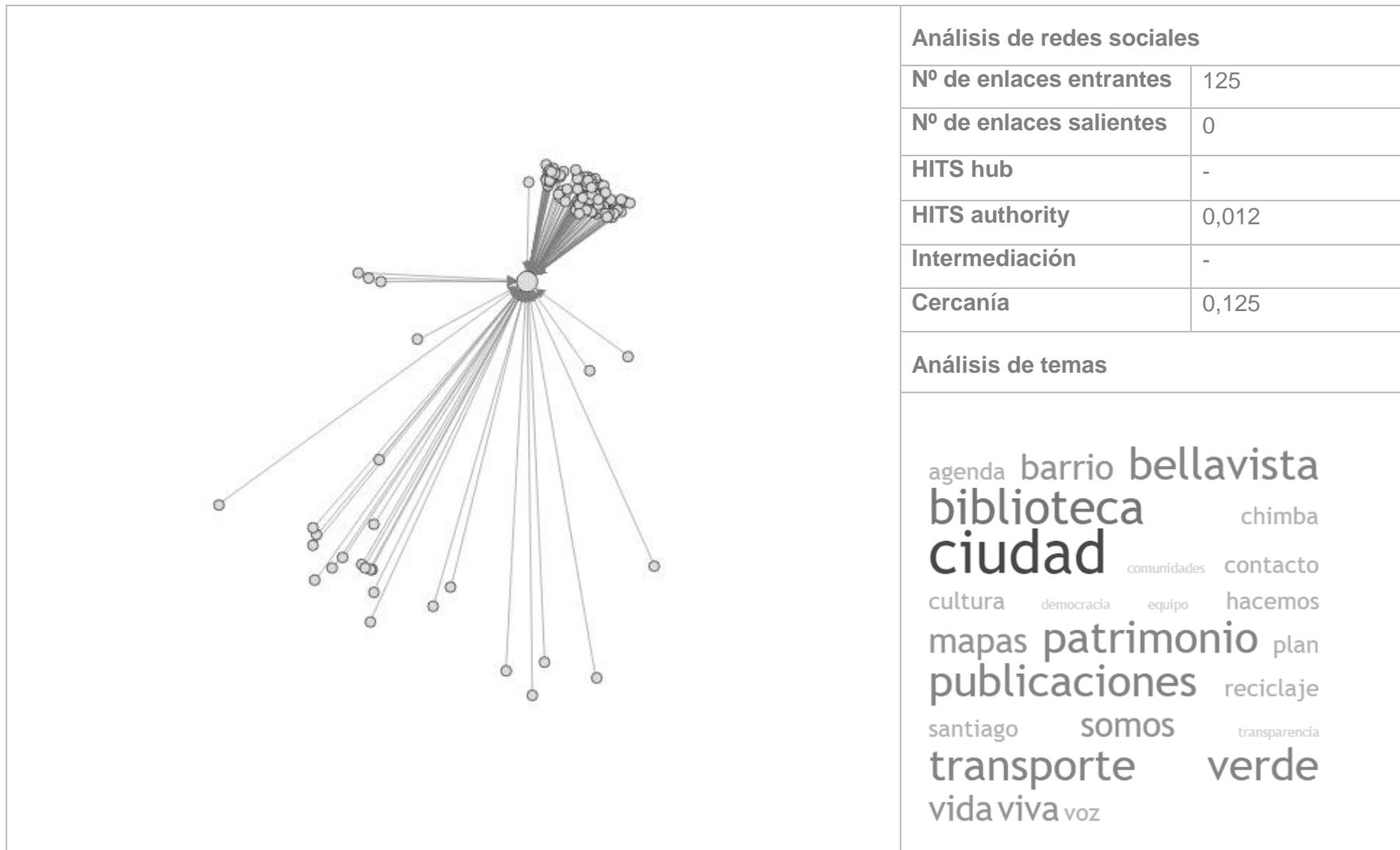


Figura 75. Representación gráfica de la red de vecindario y análisis de redes sociales de Ciudad Viva

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.bcn.cl	80	9	7	16	0,06	0,05	0,18	13828,30
www.theclinic.cl	8	1	3	4	0,01	0,00	0,16	476,07
www.latercera.com	11	1	3	4	0,01	0,00	0,16	387,13
consorcioporlatransparencia.cl	26	1	5	6	0,04	0,00	0,17	1695,42
www.az.cl	3	1	2	3	0,00	0,00	0,16	377,86
www.elquintopoder.cl	14	0	6	6	0,09	0,00	0,18	0,00
www.democraciadirecta.cl	2	0	4	4	0,03	0,00	0,17	0,00
www.innovacionciudadana.cl	6	0	3	3	0,05	0,00	0,17	0,00
www.elciudadano.cl	17	0	6	6	0,05	0,00	0,17	0,00
www.losrioscomovamos.cl	2	0	2	2	0,03	0,00	0,16	0,00
www.gugaempresas.cl	5	0	5	5	0,03	0,00	0,17	0,00
www.manzanamecanica.org	10	0	5	5	0,03	0,00	0,17	0,00
www.bicicultura.cl	5	0	2	2	0,01	0,00	0,13	0,00
www.sushiknights.cl	6	0	3	3	0,02	0,00	0,16	0,00
www.plataformaurbana.cl	37	0	3	3	0,01	0,00	0,16	0,00
www.agroecosistemas.cl	23	0	2	2	0,01	0,00	0,15	0,00
www.bellavistaenaccion.cl	27	0	2	2	0,01	0,00	0,15	0,00
www.elamaule.cl	3	0	2	2	0,01	0,00	0,15	0,00
www.solteros.cl	10	0	4	4	0,02	0,00	0,16	0,00
www.acuerdos.cl	3	0	2	2	0,01	0,00	0,15	0,00
www.avina.net	4	0	3	3	0,01	0,00	0,16	0,00
www.ciudademergente.org	4	0	2	2	0,01	0,00	0,15	0,00
www.casadelapaz.cl	2	0	3	3	0,01	0,00	0,15	0,00
www.radiotierra.cl	10	0	2	2	0,01	0,00	0,16	0,00
www.plataformaarquitectura.cl	2	0	2	2	0,01	0,00	0,16	0,00
radiotierra.info	5	0	2	2	0,01	0,00	0,16	0,00

Tabla 35. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente

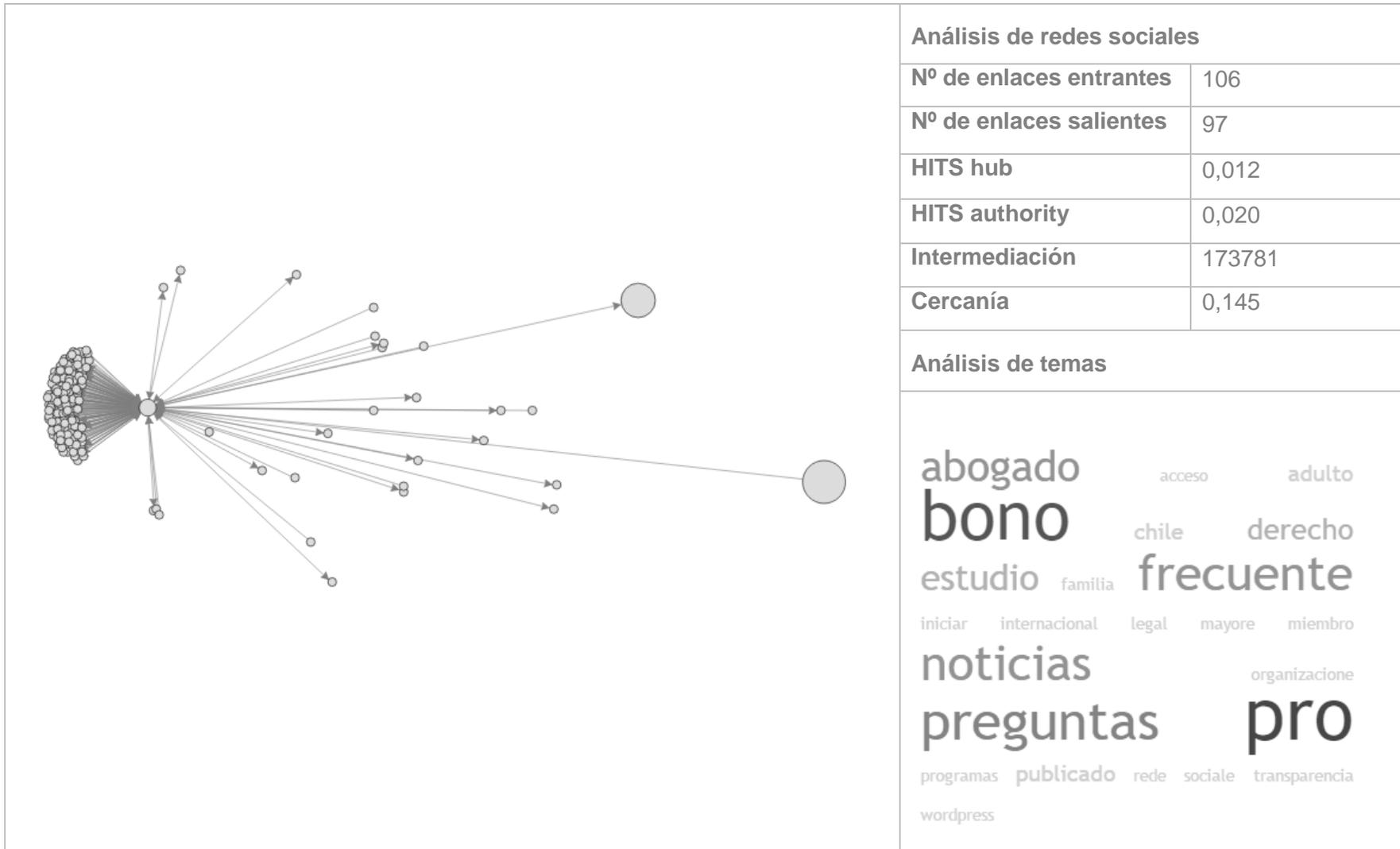


Figura 76. Representación gráfica de la red de vecindario y análisis de redes sociales de Probono

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.consejotransparencia.cl	432	108	16	124	0,06	0,38	0,17	15307,40
www.proacceso.cl	22	26	2	28	0,00	0,04	0,16	2381,42
www.twitter.com	10	5	0	5	0,00	0,01	0,13	0,00
creativecommons.org	4	5	0	5	0,00	0,02	0,13	0,00
www.youtube.com	4	3	2	5	0,03	0,01	0,16	2184,17
www.comisiondefensoraciudadana.cl	1	3	1	4	0,04	0,01	0,17	584,60
www.sernam.cl	7	2	0	2	0,00	0,01	0,13	0,00
www.flickr.com	2	2	0	2	0,00	0,01	0,13	0,00
www.walmartchile.cl	2	2	1	3	0,00	0,00	0,15	147,50
www.facebook.com	3	2	0	2	0,00	0,00	0,13	0,00
www.redprobono.org	26	1	1	2	0,00	0,00	0,16	0,00
www.az.cl	3	1	2	3	0,00	0,00	0,16	377,86
www.gqmc.cl	4	1	1	2	0,00	0,00	0,16	0,00
www.agycia.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.guerrero.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.fermandois.cl	26	1	1	2	0,00	0,00	0,16	0,00
www.palma.cl	25	1	1	2	0,00	0,00	0,16	0,00
www.alessandri.cl	4	1	1	2	0,00	0,00	0,16	0,00
www.rcz.cl	4	1	1	2	0,00	0,00	0,16	0,00
www.moralesybesa.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.pfeffer.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.bye.cl	2	1	1	2	0,00	0,00	0,16	0,00
www.gutierrezabogados.cl	18	1	1	2	0,00	0,00	0,16	0,00
www.bbis.cl	2	1	1	2	0,00	0,00	0,16	0,00
www.garciamagliona.cl	2	1	1	2	0,00	0,00	0,16	0,00
www.aranis-espinoza.cl	10	1	1	2	0,00	0,00	0,16	0,00

Tabla 36. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Probono

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.cdc.gob.cl	57	136	22	158	0,14	0,58	0,18	28946,30
www.bcn.cl	80	9	7	16	0,06	0,05	0,18	13828,30
consorcioporlatransparencia.cl	26	1	5	6	0,04	0,00	0,17	1695,42
www.redprobono.org	26	1	1	2	0,00	0,00	0,16	0,00
www.az.cl	3	1	2	3	0,00	0,00	0,16	377,86
www.gqmc.cl	4	1	1	2	0,00	0,00	0,16	0,00
www.agycia.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.guerrero.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.fermandois.cl	26	1	1	2	0,00	0,00	0,16	0,00
www.palma.cl	25	1	1	2	0,00	0,00	0,16	0,00
www.alessandri.cl	4	1	1	2	0,00	0,00	0,16	0,00
www.rcz.cl	4	1	1	2	0,00	0,00	0,16	0,00
www.moralesybesa.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.pfeffer.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.bye.cl	2	1	1	2	0,00	0,00	0,16	0,00
www.gutierrezabogados.cl	18	1	1	2	0,00	0,00	0,16	0,00
www.bbis.cl	2	1	1	2	0,00	0,00	0,16	0,00
www.garciamagliona.cl	2	1	1	2	0,00	0,00	0,16	0,00
www.aranis-espinosa.cl	10	1	1	2	0,00	0,00	0,16	0,00
www.cdrabogados.cl	5	1	1	2	0,00	0,00	0,16	0,00
www.horvitz.cl	6	1	1	2	0,00	0,00	0,16	0,00
www.vancecenter.org	6	1	1	2	0,00	0,00	0,16	0,00
www.sernac.cl	4	1	2	3	0,04	0,00	0,16	223,33
www.simondecirene.cl	1	1	1	2	0,00	0,00	0,16	0,00
www.mediale.cl	3	1	1	2	0,00	0,00	0,16	0,00
www.digitalesporchile.org	15	1	1	2	0,00	0,00	0,16	0,00

Tabla 37. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente

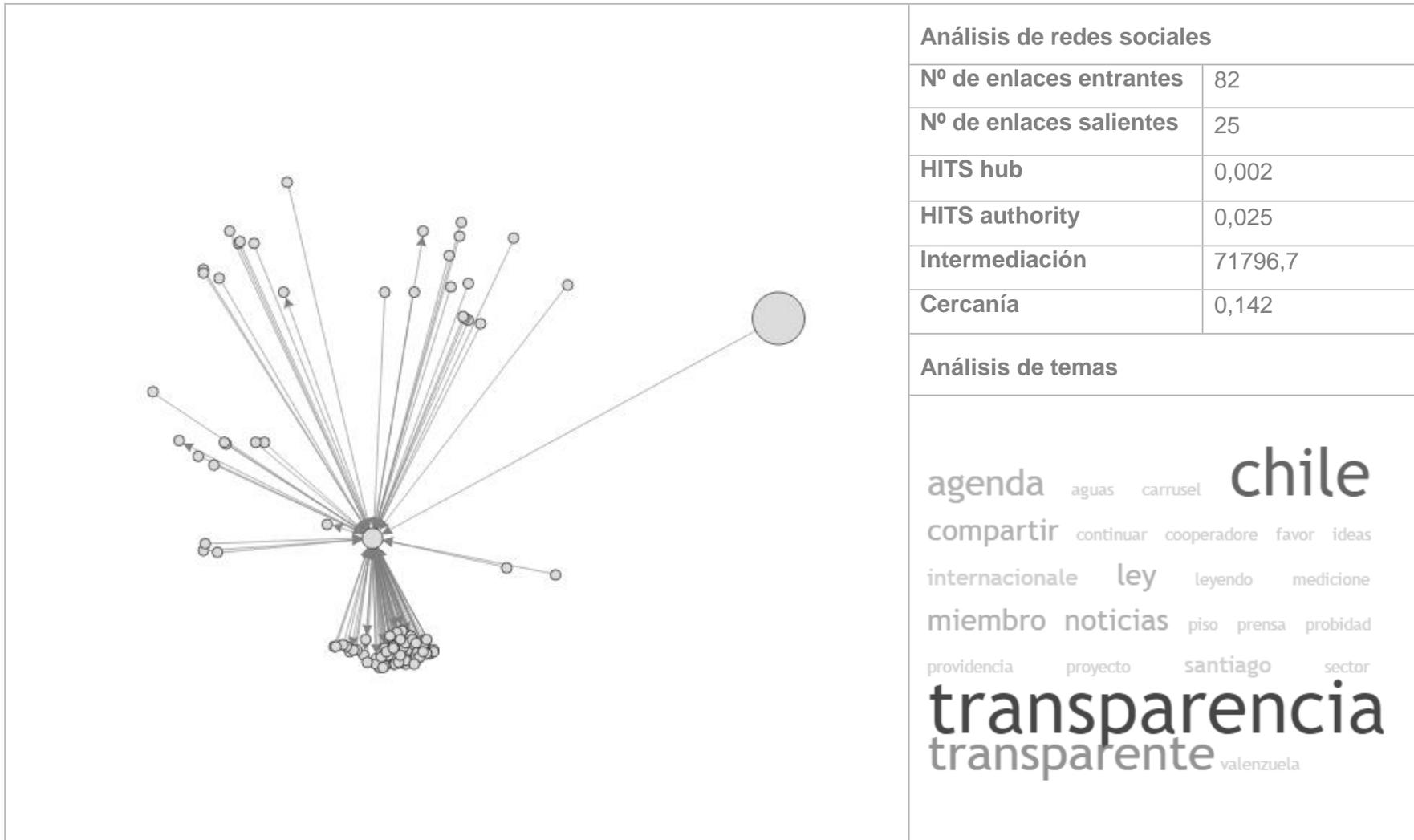


Figura 77. Representación gráfica de la red de vecindario y análisis de redes sociales de Chile Transparente

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.bcn.cl	9	7	80	16	0,06	0,05	0,18	13828,30
www.chilecompra.cl	1	1	2	2	0,03	0,00	0,16	114,75
www.transparency.org	1	1	10	2	0,00	0,00	0,15	0,00
www.partidostransparentes.cl	1	1	3	2	0,00	0,00	0,15	0,00
www.programamujer.cl	1	1	25	2	0,00	0,00	0,15	0,00
www.walmartchile.cl	2	1	2	3	0,00	0,00	0,15	147,50

Tabla 38. Identificación y análisis de redes sociales de los actores más relevantes con enlaces entrantes desde Probono

Nombre	Nº Páginas	Enlaces Entrantes	Enlaces Salientes	Total Enlaces	HITS Hub	HITS Auth	Cercanía	Intermediación
www.cdc.gob.cl	57	136	22	158	0,14	0,58	0,18	28946,30
www.bcn.cl	80	9	7	16	0,06	0,05	0,18	13828,30
www.contraloria.cl	26	2	3	5	0,06	0,01	0,16	275,52
www.walmartchile.cl	2	2	1	3	0,00	0,00	0,15	147,50
www.icei.uchile.cl	2	1	1	2	0,00	0,00	0,15	338,13
www.transparencialegislativa.org	3	1	2	3	0,01	0,00	0,16	417,02
quientefinancia.cl	3	1	2	3	0,01	0,00	0,16	575,77
leydelobby.cl	6	1	2	3	0,01	0,00	0,16	575,77
www.transparency.org	10	1	1	2	0,00	0,00	0,15	0,00
www.partidostransparentes.cl	3	1	1	2	0,00	0,00	0,15	0,00
www.programamujer.cl	25	1	1	2	0,00	0,00	0,15	0,00
www.democraciadirecta.cl	2	0	4	4	0,03	0,00	0,17	0,00
paisdigital.org	2	0	2	2	0,03	0,00	0,17	0,00
1.municipalidadalgarrobo.cl	43	0	2	2	0,03	0,00	0,16	0,00
www.right2info.org	10	0	3	3	0,03	0,00	0,16	0,00

Tabla 39. Identificación y análisis de redes sociales de los actores más relevantes con enlaces salientes hacia Gobierno Transparente

4.2.3. Análisis de temas

Después de identificar los actores y comunidades que forman parte de la red hemos procedido a identificar las diferentes posiciones, subtemas o marcos de referencia (*frames*) existentes en ella. Para ello, en primer lugar hemos analizado las informaciones de contexto relacionadas con las comunidades. Posteriormente hemos realizado un ranking de los actores basándonos en el número de enlaces entrantes que reciben y hemos analizado las informaciones de contexto de los diez actores que se pueden considerar más importantes.

La razón para analizar “únicamente” los diez actores más importantes es que este número nos permite mostrar que la metodología propuesta permite identificar diferencias en los marcos de referencia que defienden los actores en la red.

En un análisis completo de la red deberíamos analizar todos los actores que consideremos importantes pero en nuestro caso consideramos que este número nos permite alcanzar el objetivo que nos habíamos planteado que es probar la posibilidad de analizar el marco de referencia que presentan diferentes partes o comunidades de la red y más en detalle el marco de referencia de uno o varios actores en concreto.

Antes de analizar los textos recuperados hemos eliminando las palabras vacías (*stop words*) y otros términos no relevantes relacionados en su mayoría con el código fuente de las páginas. Una vez depurada la información extraída de las páginas hemos desarrollado un análisis de frecuencia de las palabras extraídas del cuerpo del texto de las páginas de inicio de los sitios que forman parte de cada clúster.

Para llevar a cabo este análisis hemos utilizado la herramienta SocSciBot en la que hemos introducido como semillas todas las URLs que forman parte de las comunidades previamente identificadas. Como proponíamos en la metodología, en el caso de las comunidades hemos extraído únicamente las páginas de inicio de cada sitio ya que ofrecen un resumen de los contenidos del resto de páginas que forman parte del sitio. Los resultados del análisis se visualizan en forma de nube de palabras asociados a las representaciones de cada clúster.

Al realizar un análisis cualitativo de los sitios más importantes de cada clúster podemos identificar rápidamente que los clústeres 1, 2, 5 y 6 están dominados por sitios del Gobierno de Chile mientras que en los clústeres 3 y 4 los principales nodos representan a organizaciones no gubernamentales.

En el caso del clúster 1 los principales nodos representan áreas de gobierno directamente relacionadas con la transparencia, el gobierno abierto y al acceso a los datos generados por el estado. La Comisión Defensora Ciudadana y Transparencia, Educatransparencia, una plataforma perteneciente al Consejo de la Transparencia que sin embargo forma parte del clúster 5, el directorio de transparencia activa Gobierno Transparente y el portal datos.gob.cl que sirve como repositorio de los datos de acceso abierto generados por el estado.

Al analizar las palabras más frecuentes en las páginas de inicio de los sitios que forman parte del Clúster 1 vemos que son Chile, gobierno y ministerio lo que resulta lógico teniendo en cuenta que los nodos que dominan este clúster representan sitios de diferentes áreas del Gobierno de Chile.

El clúster 2 está dominado por las áreas de transparencia del Ministerio de Educación, de la Dirección General de Aguas y de la Dirección General del Trabajo. Todas ellas áreas de gobierno cuya tarea principal no está relacionada con la transparencia y la probidad. El análisis de las palabras más frecuentes ofrece resultados como agua, salud o Chile. En este caso palabras relacionadas con las atribuciones de la Dirección General de Aguas y la Dirección General del Trabajo.

En el clúster 3 los principales nodos representan a organizaciones no gubernamentales dedicadas principalmente a temáticas relacionadas con la transparencia, el gobierno abierto y la participación ciudadana: Ciudadano Inteligente, Poderopedia, Ciudad Viva, Chile Transparente y Proceso. La nube de palabras de este clúster ofrece peculiaridades respecto al resto ya que presenta numerosas palabras en inglés lo que podría ser interpretado con un mayor número de relaciones internacionales con otras organizaciones.

En el clúster 4 encontramos igualmente a una organización no gubernamental, en este caso, relacionada con los servicios jurídicos dirigidos a grupos desfavorecidos y la nube de palabras asociada confirma este análisis al ofrecer palabras como abogado, ley o familia.

En el clúster 5 los principales nodos son los del Consejo de Transparencia, el Portal de Transparencia, ambos pertenecientes al Gobierno de Chile, y el portal Ley Chile de la Biblioteca del Congreso Nacional de Chile dedicado a la publicación de las leyes de Chile. El análisis de frecuencia de palabras nos ofrece en este caso una información que se confirma al realizar un análisis cualitativo de los nodos que forman parte del clúster. El nivel de administración municipal se relaciona en mayor medida con el Consejo de Transparencia y el Portal de Transparencia que con otras áreas del Gobierno de Chile.

El clúster 6 está dominado por el sitio de la Agenda de Probidad del Gobierno de Chile y la nube de palabras asociada confirma que estamos ante un proyecto del Gobierno al presentar resultados como ministerio, gobierno o estado.

Podemos afirmar pues, que el análisis de contenidos a través del análisis de frecuencia de palabras ofrece resultados que se pueden considerar lógicos en relación a los análisis cualitativos llevados a cabo sobre los sitios que forman parte de los clústers. Por lo tanto, podemos afirmar que es posible obtener información sobre los marcos de referencia utilizados por las comunidades de una red a través de un análisis automático de contenidos y evitar así el análisis cualitativo de los sitios uno a uno y reducir los recursos de tiempo que ello conllevaría.

Después de analizar los marcos de referencia de las comunidades hemos desarrollado el mismo tipo de análisis sobre los diez actores que consideramos más importantes en la red. Para definir cuáles son los actores más importantes hemos creado un ranking en base al número de enlaces entrantes de cada actor.

Para conseguir que el análisis de los marcos de referencia de los actores sea más preciso que el realizado previamente con las comunidades, hemos analizado todas las páginas de los sitios web y no únicamente las páginas de inicio de cada sitio. Los resultados del

análisis se visualizan en forma de nube de palabras asociados a las representaciones gráficas de los clústeres y de las redes de vecindario de cada actor.

En el caso del análisis de actores hemos recuperado hasta un máximo de 1000 páginas de cada sitio para llevar a cabo el análisis de frecuencia de palabras. En algunos casos este análisis no ha sido posible por restricciones impuestas por los sitios en el rastreo de sus enlaces. Ciudadano Inteligente y Ciudad Viva parecen tener restricciones en sus sitios web que no permiten llevar a cabo un análisis de sus enlaces salientes. Ni VOSON ni SocSciBot han sido capaces de recuperar los enlaces salientes de estos sitios o de recuperar las páginas que los conforman. Por ello, en estos casos hemos llevado a cabo un análisis de la página de inicio de ambos sitios, para lo que hemos utilizado la herramienta que también hemos utilizado en la creación de las nubes de palabras: Tag Crowd.

Los resultados obtenidos en los análisis de frecuencia de palabras de cada actor no presentan por tanto los mismos niveles de confiabilidad pero consideremos que sí pueden considerarse representativos del contenido del sitio ya que normalmente la página de inicio presenta una descripción de los objetivos principales de la organización y de las secciones de que se compone el sitio.

Entre los resultados del análisis de contenido de los actores encontramos que los sitios de la Comisión Defensora Ciudadana y Transparencia, de Educatransparencia, el Consejo de Transparencia y Chile Transparente son los que presentan mayor número de veces la palabra transparencia. Esto se podría interpretar como que estos actores tienen una actividad más directamente relacionada con el concepto en cuestión y no con otro tipo de actividades relacionadas como el gobierno abierto o la participación ciudadana.

Otros sitios como Ley Chile o Pro Bono presentan con más frecuencia palabras relacionadas con una perspectiva legal como ley o abogado, lo que puede interpretarse con un acercamiento de estas organizaciones al concepto de la transparencia desde un marco de referencia legal. Actores como Ciudadano Inteligente o la Comisión Defensora Ciudadana y Transparencia presentan en mayor medida palabras como ciudadano o política, más relacionadas con una perspectiva política que entiende la transparencia como un derecho de los ciudadanos.

En el caso del análisis temático de los actores no pretendemos analizar en detalle cada uno de ellos, sino tan solo mostrar que igual que sucede con los clústeres es posible extraer información relevante con un análisis automático de frecuencia de palabras sin analizar cualitativamente el sitio de cada actor en profundidad.

En el caso de los actores, a diferencia de lo que ocurre con las comunidades, no nos atrevemos a afirmar que el análisis automático de frecuencia de palabras sea más ágil y eficaz que el análisis cualitativo manual de los sitios que representan a los actores. El análisis de contenido de todas las páginas de un sitio requiere de recursos de tiempo tanto o más altos de los que requiere un análisis de un sitio llevado a cabo por una persona. Sí puede ser considerado eficaz cuando queremos analizar y representar de forma automática los contenidos de un sitio y los recursos de tiempo son menos valiosos que los recursos humanos que supondría un análisis manual.

Suponemos, por ejemplo, que queremos ofrecer un sitio web con información de diferentes organizaciones o actores relacionados con una red temática. En este caso, la metodología propuesta podría considerarse adecuada. En el caso de querer analizar una red temática de forma continuada en el tiempo y cuyos resultados deban ser actualizados frecuentemente es posible que un análisis manual resulte más eficaz.

5. DISCUSIÓN DE RESULTADOS, CONCLUSIONES E INVESTIGACIONES FUTURAS

Para discutir los resultados de la investigación retomaremos los ejes y las preguntas de investigación (PI) presentadas en la Introducción. Posteriormente presentaremos las conclusiones que extraemos de todo el proceso de investigación y finalmente propondremos algunos posibles desarrollos o investigaciones futuras.

5.1. Discusión de resultados

5.1.1. Eje de la Definición y Caracterización de Componentes de una Red Temática.

a) Caracterización de una red temática.

PG1: ¿Es posible definir de forma no ambigua y operativa a efectos de análisis e investigación el concepto de red temática en el ámbito de la web?

En el ámbito de la web, una red temática es un conjunto de individuos u organizaciones que relacionados a través de una red de hiperenlaces plantean opiniones y utilizan recursos u oportunidades para influir en un tema de discusión. Esta primera propuesta, que en su momento fue elaborada como una propuesta de definición nos ha permitido el modelado de sus componentes (como se argumenta en la Conclusión 2) y ha demostrado ser lo suficientemente operativa como para ser adoptada en nuestra investigación.

b) Componentes de una red temática.

PE1: ¿Es posible identificar de forma operativa los componentes principales y suficientes a efectos de análisis e investigación de una red temática en el ámbito de la web?

De acuerdo con nuestra propuesta, los tres componentes suficientes que intervienen en nuestro modelo teórico de red temática son: (1) las redes de hiperenlaces, (2) el tema o temas y (3) los actores. Como en la conclusión número 1, esta propuesta de composición de una red temática ha demostrado ser suficientemente operativa como para conducir nuestra investigación. Además, obviamente puede ser utilizada para futuras investigaciones o por parte de otros investigadores.

c) Proposiciones conceptuales.

PE2: ¿Es posible presentar un conjunto de propuestas conceptuales sobre las características de los componentes que forman parte de una red temática en el ámbito de la web?

- En ciencias sociales y en ciencias de la información, una **red de hiperenlaces** es el sistema dinámico que hace posible el flujo de información entre diferentes sitios web y a través del cual se establecen relaciones de comunidad entre los actores que los crean.
- Un **tema** es un conjunto de ideas, teorías o conceptos que generan controversia entre grupos de actores políticos, sociales o económicos.
- Un **actor** es todo individuo u organización, del ámbito público o privado, que tenga quejas, recursos u oportunidad de influir en un tema de discusión.

5.1.2. Eje de la Caracterización y Evaluación de Herramientas

PG2: ¿Cuáles son las herramientas que podemos considerar más adecuadas para el análisis de redes temáticas en la web?

PE1: ¿Cuáles son las características de las herramientas que se pueden considerar más adecuadas para el análisis de redes temáticas en la web?

PE2: ¿Cuáles son los parámetros e indicadores que permiten evaluar las herramientas relacionadas con el análisis de redes temáticas en la web?

Dadas sus especiales características las conclusiones relacionadas con las preguntas del eje de la Caracterización y Evaluación de Herramientas se presentan condensadas en la Figura 53 y 54 que presentamos en la siguiente página.

5.1.3. Eje del Sistema de Análisis de Redes Temáticas en la Web

PG3: ¿Podemos desarrollar un sistema de análisis suficientemente escalable y ágil para adaptarse a la dinámica característica de los fenómenos de opinión que tienen lugar en la web?

a) Rendimiento y potencialidad del Sistema de Análisis.

Los resultados obtenidos en el análisis de la Red Temática de la Transparencia en Chile nos permiten afirmar que el sistema de análisis propuesto ofrece buenos resultados y a la vez es ágil y fácil de implementar. Desde un punto de vista estrictamente académico o conceptual, y por tanto, ignorando aspectos contractuales (tarifas de precios), podemos decir que además **es escalable** para permitir el análisis de un mayor número y tamaño de redes.

Atributos	Indicadores	IssueCrawler	SocSciBot	Webometric Analyst	VOSON
Eficiencia	Comportamiento temporal	En función de la demanda	Variable en función del ancho de banda	Variable en función del ancho de banda	En función de la demanda y del tipo de suscripción
	Utilización de recursos	Procesamiento en la nube	Bajo consumo de recursos. Alto consumo de ancho de banda.	Bajo consumo de recursos. Alto consumo de ancho de banda.	Procesamiento en la nube
	Capacidad	Limitada	Ilimitada	Limitada gratuita – Ilimitada de pago	Limitada al tipo de suscripción – Ilimitada de pago
Compatibilidad	Coexistencia	Todos los navegadores	S.O. Windows	S.O. Windows	Firefox – Suscripción gratuita Todos los navegadores – Otras suscripciones
	Exportación	.svg, .jpg, .png, .tiff, .pdf, .html, .net, .xml, .txt, .gefx	.net, .txt	.net, .jpg, .txt	.png, .svg, .gml, .net, .csv
Usabilidad	Documentación	Regular	Regular	Regular	Buena
	Aprendizaje	Fácil	Medio	Difícil	Medio
	Operabilidad	Alta	Media	Media	Alta
	Estética	Buena	Mala	Regular	Muy buena
	Accesibilidad	Regular	Mala	Mala	Regular
Seguridad	Confidencialidad	Mala	Regular	Regular	Buena
	Responsabilidad	-	Alta	Alta	Media
Portabilidad	Instalación	-	Media	Fácil	-
	Coexistencia	-	Media	Media	-
	Reemplazo	-	Medio	Fácil	-

Figura 53. Evaluación de la calidad general en uso de IssueCrawler, SocSciBot, Webometric Analyst y VOSON

Atributos	Indicadores	IssueCrawler	SocSciBot	Webometric Analyst	VOSON
Análisis de hiperenlaces	Enlaces entrantes	x	x	✓	✓
	Enlaces salientes	✓	✓	✓	✓
	Coenlaces salientes	✓	x	x	✓
	Coenlaces entrantes	x	x	✓	✓
	Análisis estadístico (SNA)	✓	✓	✓	✓
	Opciones de preprocesamiento	x	Archivos de texto	Archivos de texto	Base de datos
Análisis de contenidos	Palabras clave	x	✓	x	✓
	Textos completos	x	✓	x	✓
	Opciones de preprocesamiento	x	✓	✓	✓
	Búsqueda de palabras	x	✓	x	✓
	Análisis de frecuencia de palabras	x	✓		✓
	Análisis vectorial de contenidos	x	x	x	✓
Visualización de redes	Algoritmos de representación	Indegree centrality	Fruchterman-Reingold	- Fruchterman-Reingold - Kamada Kawai	- Fruchterman-Reingold - LinLog de Noack
	Configuración de nodos	✓	✓	✓	✓
	Configuración de enlaces	✓	✓	✓	✓
	Opciones de exportación	.svg, .jpg, .png, .tiff, .pdf, .html, .net, .xml, .txt, .gefx	.net, .txt	.net, .jpg, .txt	.png, .svg, .gml, .net, .csv

Figura 54. Evaluación de la calidad funcional de IssueCrawler, SocSciBot, Webometric Analyst y VOSON

b) Capacidad de identificación

La metodología propuesta para el análisis de redes de hiperenlaces ha ofrecido los resultados esperados permitiendo identificar de manera ágil los sitios relacionados con un tema de discusión y ofreciendo representaciones significativas de la red temática tanto en su versión absoluta como en su versión depurada.

c) Economía de medios

La metodología propuesta para el análisis de actores presenta la dificultad del necesario procesamiento de nodos para identificar con exactitud a los actores. A pesar de esto hemos comprobado que es posible identificar y analizar los principales actores de una red sin llevar a cabo este paso previo. El procesamiento previo de los nodos puede ser de gran valor en el análisis detallado de una red pero no nos parece indispensable para poder analizar las relaciones que mantiene un actor y las posiciones que defiende.

d) Fases y Componentes de un Sistema de Análisis de Redes Temática

PE1: ¿Qué fases y componentes puede tener un sistema de análisis escalable para analizar redes temáticas en la web?

Hemos desarrollado, presentado y puesto a prueba un sistema de análisis. En particular, el sistema de análisis que defendemos se divide en tres fases que se corresponden con los que consideramos tres componentes de las redes temáticas en la web.

- a) **Análisis de redes de hiperenlaces:** consiste en identificar los sitios web más relevantes relacionados con un tema de discusión y las relaciones que se construyen entre ellos a través del uso de hiperenlaces.

- b) **Análisis de temas:** consiste en identificar los diferentes subtemas o marcos de referencia existentes en una red temática a través del análisis de contenido de los textos de los sitios que forman la red.

- c) **Análisis de actores:** consiste en identificar los actores que forman parte de la web, su papel en la estructura de la red y las posturas que defienden respecto al tema de discusión.

e) Análisis de Redes y Periodismo de Datos

A través del desarrollo de un análisis-test de una Red Temática, la Red Temática de la Transparencia en Chile, hemos presentado un caso que ilustra las posibilidades de nuestra investigación en general, y del sistema de análisis en particular en el ámbito de la Comunicación. Por un lado, el Periodismo de Datos podría utilizar el sistema de análisis presentado aquí, o alguna variación basada en el mismo. Además de permitir un detallado análisis de los actores, temas e influencias recíprocas en redes temáticas, el sistema facilita su representación mediante diferentes formatos de representación visual de la información.

f) Fases de Aplicación del Sistema de Análisis

PE2: ¿Cómo podrían operacionalizarse las tres fases del análisis de redes temáticas en la web: análisis de redes de hiperenlaces, análisis de temas y análisis de actores?

A continuación presentamos la metodología concreta propuesta para llevar a cabo las tres fases del análisis: análisis de hiperenlaces, análisis de temas y análisis de actores.

Fases y componentes del análisis de redes de hiperenlaces

Selección de semillas

- Consultas en buscadores
- Directorios
- Medios de comunicación
- Consulta de expertos
- Rastreo de co-enlaces entrantes

Análisis de enlaces

- Enlaces entrantes
- Enlaces salientes
- Enlaces recíprocos
- Coenlaces entrantes
- Coenlaces salientes

Análisis de redes sociales

- Clústering
- Densidad
- Centralidad
- Intermediación
- Cercanía

Visualización de redes

- Red absoluta
- Red de coenlaces
- Clústers

Fases y componentes del análisis de actores

Identificación de actores

- Agrupamiento de resultados
- Desagrupamiento de resultados

Análisis de redes sociales

- Centralidad
- Intermediación
- Cercanía
- HITS

Visualización de redes de vecindario

Fases y componentes del análisis de temas.

Análisis de comunidades

- Frecuencia de palabras en las páginas de inicio del sitio

Análisis de actores

- Frecuencia de palabras en todas las páginas del sitio

Visualización de nubes de palabras

- Palabras asociadas a las comunidades
- Palabras asociadas a los actores

5.1.4. Eje del Estudio de Caso de la Red Temática de la Transparencia en Chile

a) Aplicación al análisis de fenómenos de opinión

PG4: ¿Es posible aplicar el concepto de redes temáticas al estudio de los fenómenos de opinión que tienen lugar en la web?

Los resultados obtenidos en el análisis de la Red Temática de la Transparencia en Chile nos permiten afirmar que la aplicación del concepto de red temática mediante un sistema de análisis adecuado puede ofrecer información valiosa para la comprensión de los fenómenos de opinión que se dan en la red.

b) Caracterización de una red de hiperenlaces

PE1: ¿Es posible describir las características de la red de hiperenlaces que subyace a la Red Temática de la Transparencia en Chile?

Los resultados obtenidos en el análisis de la red de hiperenlaces que subyace a la Red Temática de la Transparencia en Chile nos permiten afirmar que la aplicación de las medidas de análisis de redes sociales utilizadas, es decir, número de nodos, número de enlaces, densidad, centralidad, intermediación y cercanía, ofrecen información relevante sobre las características de la red y sobre sus dinámicas internas. Sin embargo, el potencial de este tipo de análisis se demuestra en mayor medida en la comparación de dos o más redes y pierde relevancia en el caso de analizar una única red.

c) Identificación de actores o comunidades

PE2: ¿Es posible identificar las comunidades existentes en la Red Temática de la Transparencia en Chile?

Los resultados obtenidos en la identificación de comunidades dentro de la Red Temática de la Transparencia en Chile nos han permitido identificar grupos de actores con características similares. Se han identificado, por ejemplo, dos comunidades cuyos actores principales pertenecen en el primer caso al Gobierno de Chile y en el segundo caso son organizaciones no gubernamentales. Esto nos permite afirmar que el algoritmo de clústering utilizado ofrece resultados relevantes que permiten obtener información sobre la naturaleza de los diferentes actores que forman la red.

d) Identificación de posiciones o marcos de referencia en una Red Temática

PE3: ¿Es posible identificar las posiciones o marcos de referencia que defienden las diferentes comunidades existentes en la Red Temática de la Transparencia en Chile?

El análisis de contenido a partir del análisis de frecuencia de palabras nos ha permitido extraer información relevante relacionada con el marco de referencia que utiliza cada comunidad. No obstante, consideramos que otros tipos de análisis de contenido automático como el análisis vectorial o el agrupamiento de textos en clústers podrían ofrecer resultados más precisos e información más relevante sobre las comunidades.

e) Identificación de actores relevantes en una red temática

PE4: ¿Es posible identificar los actores más relevantes relacionados con el tema de la transparencia en Chile?

El análisis de redes sociales aplicado sobre los actores de la red ofrece resultados que permiten identificar los actores más relevantes en función de diferentes criterios. En esta investigación hemos utilizado la centralidad de un actor en la red a partir de los enlaces entrantes que recibe como criterio principal de relevancia. Sin embargo, otras medidas de análisis como la intermediación o la cercanía nos permiten igualmente identificar los actores más relevantes en función de otros criterios en función de las necesidades del análisis. Por todo lo anterior consideramos que la metodología propuesta para el análisis de actores ofrece resultados que permiten alcanzar los objetivos planteados para el estudio de redes temáticas en la web.

f) Identificación de relaciones entre los actores de una red temática

PE5: ¿Es posible identificar las relaciones que mantiene un actor de la Red Temática de la Transparencia en Chile con el resto de actores de la red?

La identificación de las relaciones de un actor con el resto de la red es posible a través de la utilización de herramientas de análisis que permitan el análisis de enlaces entrantes y salientes de un sitio web. Como explicábamos en el capítulo 3 dedicado a la descripción y evaluación de las herramientas de análisis, actualmente los buscadores han limitado el acceso a la información de sus crawlers, lo que hace difícil obtener información sobre los enlaces entrantes de un sitio web. Webometric Analyst ofrece la posibilidad de conseguir esta información a través del servicio de pago Microsoft Market Azure y VOSON desarrolla su análisis de enlaces entrantes a partir de la información del crawler SMOZ también previo pago de una suscripción. Por lo tanto, podemos afirmar que sí es posible identificar las relaciones que mantienen los actores de una red temática en la web pero ello tiene un coste monetario asociado que hace difícil la escalabilidad de la metodología propuesta.

g) Identificación de posturas de los actores en una red temática

PE6: ¿Es posible identificar las posiciones que defienden los diferentes actores de la Red Temática de la Transparencia en Chile?

El análisis de contenido a partir del análisis de frecuencia de palabras nos ha permitido extraer información relevante relacionada con el marco de referencia que utiliza cada actor. No obstante, y como señalábamos en el caso del análisis de contenido de los clústers, consideramos que otros tipos de análisis de contenido automático como el análisis vectorial o el agrupamiento de textos en clústers podrían ofrecer resultados más precisos e información más relevante sobre los marcos de referencia de cada actor.

5.2. Conclusiones

Consideramos que en esta investigación en su conjunto ofrece una aproximación novedosa a los estudios de los efectos de la comunicación y a otros ámbitos de investigación en ciencias sociales que requieren o pueden servirse de la información que contiene la web. El sistema de análisis propuesto, además de permitir un detallado análisis de las redes temáticas en la web, facilita su representación mediante diferentes formatos de representación visual, lo que creemos podría ser aprovechado en el ámbito del Periodismo de Datos o en cualquier ámbito de conocimiento que requiera de técnicas de análisis y representación de la información. Entendemos que facilitar este tipo de acceso y representación de la información tiene en sí mismo una relevancia política puesto que favorece el conocimiento y la transparencia de fenómenos de formación de la opinión pública.

Analizando las diferentes fases y objetivos de esta investigación, consideramos que el concepto de red temática que proponemos abre nuevas vías para la investigación, no solo de los procesos de formación de la opinión pública en la web, sino que ayuda al desarrollo del ámbito de la comunicación digital y propone nuevas líneas de investigación en Comunicación Social. Entendemos que los conocimientos aportados en esta caracterización de componentes de las redes temáticas pueden ser útiles, siempre con las

adaptaciones que convengan en cada caso, a investigaciones relativas tanto al ámbito de la sociología, las ciencias políticas como las ciencias empresariales.

El diseño y desarrollo de un sistema de análisis de redes temáticas en la web ha supuesto integrar un conjunto de técnicas de recuperación, análisis y representación de la información que aunque ya habían sido previamente utilizadas por separado en el ámbito de la investigación científica, pensamos que no habían sido suficientemente sistematizadas en su uso combinado. A través de estas técnicas ha sido posible operacionalizar el desarrollo teórico del concepto de red temática y sus componentes en un conjunto de fases y técnicas de análisis que pueden ser aplicables al ámbitos de estudio diversos como los fenómenos de opinión pública, los movimientos sociales, los estudios de mercado o los procesos de comunicación política en la web puesto que en todos ellos existen y es necesario analizar la actividad de grupos de actores que comparten un ámbito temático de interés común y establecen relaciones a través de la tecnología digital del hipertexto.

Hemos analizado, descrito y evaluado las herramientas crawler más relevantes en el ámbito de la investigación científica, VOSON, IssueCrawler, SocSciBot y Webometric Analyst, lo que consideramos puede ser valioso para cualquier investigador que requiera llevar a cabo investigaciones sobre los hipertextos que se producen en la web en general o incluso en plataformas sociales de comunicación.

El estudio de un caso que hemos realizado, a pesar de que tiene un valor principalmente instrumental en nuestra investigación y nos sirve para validar el sistema de análisis y el modelo teórico de análisis de redes temáticas en la web que habíamos propuesto, puede tener un valor social y político desde el momento en que aporta cierto conocimiento sobre las instituciones públicas y privadas de Chile que tienen alguna relación con el tema de discusión de la transparencia. Gracias a los resultados del estudio de la red temática de la transparencia en Chile ha sido posible generar conocimiento sobre los actores más relevantes que intervienen en el debate de la transparencia en Chile, las comunidades y relaciones que existen entre actores gubernamentales y actores del tercer sector, por ejemplo, y las diferentes perspectivas que utiliza cada actor en relación al tema de la transparencia. Consideramos que cualquier aporte en este sentido puede ser valioso para

las propias organizaciones que participan en la red y para cualquier actor interesado en formar parte de ella. De igual manera, los resultados del estudio de caso pueden ser relevantes para cualquier investigador que necesite datos sobre las organizaciones que tienen relación con la transparencia o para investigación que tengan como objeto de estudio la transparencia en general o en Chile en particular.

5.3. Investigaciones futuras

En relación a las posibles investigaciones futuras o desarrollos de esta investigación, nos planteamos tres ejes fundamentales que presentamos a continuación.

5.3.1. Análisis de plataformas sociales

En futuras investigaciones nos proponemos incorporar al sistema de análisis de redes temáticas en la web el análisis de información contenida en plataformas sociales como Twitter, LinkedIn, Mendeley o Research Gate.

En esta investigación nos hemos centrado en el análisis de redes temáticas que establecían su comunicación a través de hiperenlaces en sus sitios web, sin embargo, somos conscientes de que actualmente un porcentaje importante de la comunicación entre instituciones tiene lugar en las redes sociales por lo que para entender con mayor profundidad los procesos de formación de opinión y la evolución de los debates en torno a temas de controversia social, consideramos que es necesario analizar la comunicación que se produce a través de este tipo de herramientas y aportar así datos más completos sobre las dinámicas de las redes temáticas en la web.

Al analizar las herramientas de análisis de hiperenlaces, hemos visto que Webometric Analyst incorpora utilidades que permiten analizar, a día de hoy pues como ya hemos visto las condiciones de este tipo de servicios para acceder a su información puede cambiar en cualquier momento, plataformas sociales de comunicación como Youtube, Flickr, Twitter, Mendeley, Research Gate, Academia o Tumblr, por lo que creemos realista plantear la posibilidad de llevar a cabo investigaciones que analicen información extraída de este tipo de plataformas.

5.3.2. Análisis de contenido

En relación al análisis de temas, consideramos necesario incorporar herramientas especializadas de análisis de contenido que permitan desarrollar análisis más sofisticados que el análisis de frecuencia de palabras que hemos utilizado en esta investigación. Estimamos que es necesario emplear herramientas que ofrezcan funcionalidades más avanzadas de preprocesamiento de los textos.

El análisis de contenido a partir del análisis de frecuencia de palabras nos ha permitido extraer información relevante relacionada con el marco de referencia que utiliza cada comunidad y actores pero entendemos que otros tipos de análisis de contenido automático como el análisis vectorial o el agrupamiento de textos en clústers podrían ofrecer resultados más precisos e información más relevante sobre los textos extraídos.

El análisis de contenido puede realizarse a través de métodos supervisados de aprendizaje (*supervised learning*) como Naïve Bayes, y existen herramientas como Rapidminer que hacen relativamente asequible para cualquier investigador desarrollar este tipo de análisis. Existen además otras alternativas como las que ofrece el proyecto Clarin que pone a disposición de cualquier investigador un conjunto de herramientas de corpus y análisis lingüístico que tiene la misión de promover y asesorar en el uso de tecnología y herramientas de análisis de textos en la investigación en Humanidades y Ciencias Sociales.

5.3.3. Correlación entre clústers de redes y clústers de contenidos

En la investigación que presentamos hemos intentado identificar las diferentes comunidades que existen en la red temática de la transparencia en Chile. Para ello hemos utilizado el algoritmo Newman – Girvan que nos ha permitido identificar y analizar seis comunidades en la red analizada. En una investigación futura consideramos que sería valioso estudiar si el agrupamiento en clústers a partir de la aplicación del algoritmo Newman-Girvan sobre los nodos de la red ofrece resultados correlativos con los que se obtendrían a través del clústering de los textos extraídos de los sitios web de la misma

red. Es decir, si el clústering de actores a partir de sus hiperenlaces coincide con el agrupamiento de actores a partir del análisis del contenido y agrupamiento en clústers de los textos extraídos de los sitios web que forman la red.

BIBLIOGRAFÍA

- Abadal, E., & Codina, L. (2005). Recuperación de Información. *Bases de Datos Documentales: Características, funciones y método*, 29-92.
- Ackland, R. (2005). "Mapping the U.S. Political Blogosphere: Are Conservative Bloggers More Prominent?". *BlogTalk Downunder*. <http://hdl.handle.net/1885/45827>
- Ackland, R.; Gibson, R. K. (2005) "Hyperlinks and horizontal political communication on the WWW: The untold story of parties online". *Virtual Observatory for the Study of Online Networks*". <http://hdl.handle.net/1885/45827>
- Ackland, R., O'Neil, M., Standish, R., & Buchhorn, M. (2006). VOSON: A Web services approach for facilitating research into online networks. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.1055&rep=rep1&type=pdf>
- Ackland, R. et al. (2007). "Mapping 'small things' on the Web: Assessing the online presence of the nanotechnology industry". *Annual Conference of the International Communication Association*, 24-28.
- Agostini, A. (1984). "La tematizzazione. Selezione e memoria dell'informazione giornalistica". *Problemi dell'Informazione*. 9(4), 51-63.
- Al-Badareen, A. B., Selamat, M. H., Jabar, M. A., Din, J., & Turaev, S. (2011). Software Quality Models: A Comparative Study. In *Software Engineering and Computer Systems*, 46-55. Springer Berlin Heidelberg. https://www.researchgate.net/profile/Sherzod_Turaev/publication/220868681_Software_Quality_Models_A_Comparative_Study/links/0deec5190e5aa28921000000.pdf
- Almind, T. C.; Ingwersen, P. (1997) "Informetric analyses on the world wide web: methodological approaches to webometrics", *Journal of Documentation*, 53(4), 404-426.
- Badia, L. (1992). *De la persuasió a la tematizació: Introducció a la comunicació política moderna*. Barcelona: Portic.
- Barabási, A. L. (2002). *Linked: The new science of networks*. Cambridge, Massachusetts: Perseus Publishing.
- Barabási, A.L.; Albert, R. (1999). "Emergence of scaling in random networks". *Science*, 286, 509-512.
- Bavelas, A. (1948). "A mathematical model for group structure". *Applied Anthropology* 7:16-30.
- Berrio, J. (2002). "L'Opinió Pública". *Portal de la Comunicació Incom-UAB*. http://www.portalcomunicacion.com/ESP/pdf/aab_lec/5.pdf
- Berry, J. M. (1984). *The interest group society*. Boston: Little Brown.

- Björneborn, L.; Ingwersen, P. (2004). "Toward a basic framework for webometrics". *Journal of the American Society for Information Science and Technology*, 55(14), 1216-1227.
- Blázquez Ochando, M. (2014). Nuevos retos de la tecnología web crawler para la recuperación de información. *Métodos de información*, 4(7), 115-128.
- Blumler, J. G. (2001). "The third age of political communication". *Journal of public affairs*, 1(3), 201-209.
- Boehm, B. W.; Brown, J. R.; Kaspar, H. (1978). Characteristics of software quality. *TRW series of software technology*, 1.
- Breu, R.; Kuntzmann-Combelles, A.; Felderer, M. (2014). "New perspectives on software quality". *IEEE Software*, (1), 32-38.
https://www.researchgate.net/profile/Michael_Felderer/publication/260526165_New_Perspectives_on_Software_Quality/links/543d6cb80cf25d6b1ad94978.pdf
- Brin, S.; Page, L. (1998). "The Anatomy of a Large-Scale Hypertextual Web Search Engine". *Computer Networks and ISDN Systems*, 30(1-7), 107-117.
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.517.3557&rep=rep1&type=pdf>
- Broder, A. et al. (2000). Graph structure in the web. *Computer networks*, 33(1), 309-320.
- Castells, M. (1997). *La era de la información*, vol. 1. Madrid: Alianza Editorial.
- Castells, M. (2009). *Comunicación y poder*. Madrid: Alianza Editorial.
- Castillejo, M. (2011). "Metodología para el análisis de la opinión pública y la comunicación política en Internet" <http://hdl.handle.net/10230/11356>
- Castillejo, M.; Codina, L.; Pont, C. (2014). "Redes temáticas en la web: una propuesta de componentes y caracterización." *Hipertext.net* [online], núm. 12. <http://raco.cat/index.php/Hipertext/article/view/274307/364485>
- Chakrabarti, S.; Van den Berg, M.; Dom, B. (1999). "Focused crawling: a new approach to topic-specific Web resource discovery". *Computer Networks*, 31(11), 1623-1640.
- Chaffee, S. H.; Metzger, M. J. (2001). "The end of mass communication?". *Mass communication and society*, 4(4), 365-379
- Cho, J.; Garcia-Molina, H.; Page, L. (1998). Efficient crawling through URL ordering. Disponible en <http://ilpubs.stanford.edu:8090/347/1/1998-51.pdf>
- Cobb, R. W.; Elder, C. D., (1972). "Participation in American Politics: The Dynamics of Agenda Building". *The Journal of Politics*, 33(4), 892-915.
- Codina, L. (2000). Evaluación de recursos digitales en línea: conceptos, indicadores y métodos. *Revista española de documentación científica*, 23(1), 9-44.

Codina, L. (2010). "El futur de la informació acadèmica: ¿web semántica, web social o totes dues?" *La Terminologia i la Documentació: Relacions i Sinergies: Actes de la VII Jornada de la Scaterm*. Barcelona: Societat Catalana de Terminologia.

Codina, L. (2015). No lo llame Análisis Bibliográfico, llámelo Revisión Sistematizada. Y cómo llevarla a cabo con garantías: Systematized Reviews + SALSA Framework. Recuperado de <http://www.lluiscodina.com/revision-sistemica-salsa-framework/>

Codina, L., Pedraza, R., Noci, J. D., Rodríguez-Martínez, R., Pérez-Montoro, M., y Cavaller-Reyes, V. (2014). Sistema Articulado de Análisis de Cibermedios (SAAC): Una propuesta sobre el qué y el cómo para estudiar medios de comunicación digitales. *Hipertext. net*, (12).

Dasgupta, A. et al. (2007). The discoverability of the web. In *Proceedings of the 16th international conference on World Wide Web*, 421-430. ACM.
<http://infolab.stanford.edu/~olston/publications/discovery.pdf>

Dreze, X.; Zufryden, F. (2004). "The Measurement of online Visibility and its Impact on Internet Traffic". *Journal of Interactive Marketing*, 18(1), 20-37.

Dromey, G. R. (1995). A model for software product quality. *Software Engineering, IEEE Transactions on*, 21(2), 146-162.
http://www.itu.dk/~jacok/VIPS/brugbare/a_model_for_software_product_quality.pdf

Eeles, P. (2005). Capturing architectural requirements. *IBM Rational developer works*.
http://itculiacan.webs.com/descargas/z_FURPS%20Requisites.pdf

Eguíluz, V.M.; Toral, R.; San Miguel, M. (2005). "Redes complejas en la dinámica social". *Inguruak, Revista Vasca de Sociología y Ciencia Política*, 42, 127-146.

Erdos, P.; Reyni, A. (1960). "On the evolution of random networks". *Mathematical Institute of Hungarian Academic Science*, 5, 17-61.

Etzkowitz, H., & Leydesdorff, L. (1995). The Triple Helix--University-industry-government relations: A laboratory for knowledge based economic development. *Easst Review*, 14(1), 14-19.

Fernández-Silva, S. (2013). Variación denominativa y punto de vista. *Revista Debate Terminológico*, 9, 11-37.

Foot, K. A.; Schneider, S.M. (2002). "Online action in campaign 2000: An Exploratory Analysis of the U.S. Political Web Sphere". *Journal of Broadcasting and Electronic Media*, 46(2), 222-244.

Freeman, L. C. (1977). "A set of measures of centrality based on betweenness". *Sociometry*, 35-41.

Freeman, R. E. (1984). *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

- Freeman, L. C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver: Empirical Press.
- Frooman, J. (2010). "The issue network: reshaping the stakeholder model". *Canadian Journal of Administrative Sciences*, 27(2), 161-173.
- Fuchs, C. (2006). "The self organization of social movements". *Systemic Practice and Action Research*, 19(1), 101-137.
- Galambos, L. (1983). "Technology, Political Economy and Professionalization: Central Themes for the Organizational Synthesis," *Business History Review*, 57, 471-93.
- Grady, R. B. (1992). *Practical software metrics for project management and process improvement*. Prentice-Hall, Inc..
- Grady, R. B.; Caswell, D. L. (1987). Software metrics: establishing a company-wide program.
- Gray, M. (1993). Internet growth and statistics: Credits and background.
<http://www.mit.edu/people/mkgray/growth/>
- Hamm, K. E. (1983). Patterns of influence among committees, agencies, and interest groups. *Legislative Studies Quarterly*, 379-426.
- Hayden, F. (2002). "Policymaking Network of the Iron-Triangle Subgovernment for Licensing Hazardous Waste Facilities". *Journal of Economic Issues* 36 (2): 479. :
<http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1007&context=cbafacepub>
- Heclo, H. (1978). Issue networks and the executive establishment. *Public administration: Concepts and cases*, 413.
- Henzinger, M. (2001). "Hyperlink Analysis for the Web" *IEEE Internet Computing* 5(1): 45-50.
- Hsu, C. L.; Park, H. W. (2012). "Mapping online social networks of Korean politicians". *Government Information Quarterly*, vol. 29: núm. 2, pp. 169-181.
- ISO/IEC (1999). 14598-1:1999 Information technology: Software product evaluation
Part 1: General overview.
- ISO/IEC (2001). 9126-1, Software engineering - product quality - Part 1: Quality Model.
- ISO/IEC (2001). 9126-2, Software engineering - product quality - Part 2: External metrics.
- ISO/IEC (2001). 9126-3, Software engineering - product quality - Part 3: Internal Metrics.
- ISO/IEC (2001). 9126-4, Software engineering - product quality - Part 4: Quality In Use Metrics.

ISO/IEC (2005). 25000, Software Engineering - Software product Quality Requirements and Evaluation (SQuaRE) - Guide to SQuaRE.

ISO/IEC (2011). 25010, Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models.

IssueCrawler [software] (2001). Amsterdam: Govcom.org. Obtenido de www.issuecrawler.net

Jackson, M.H. (1997). "Assessing the structure of communication on the world wide web". *Journal of Computer-Mediated Communication*, 3(1). Disponible en <http://jcmc.indiana.edu/vol3/issue1/jackson.html>

Jenkins, J. C. (1995). *Social movements, political representation, and the state*. Jenkins, J. C. *The politics of social protest*, Minneapolis: University of Minnesota, 14-38.

Kleinberg, J.C. (1999). "Authoritative sources in a hyperlinked environment". *Journal of the ACM*, vol. 46(5), 604-642.

Koopmans, R.; Zimmermann, A. (2003). "Internet: A New Potential for European Political Communication?". *Discussion Paper SP IV 2003-402*. Wissenschaftszentrum Berlin für Sozialforschung (WZB).

Koopmans, R.; Zimmermann, A. (2007). "Visibility and communication networks on the Internet: the role of search engines and hyperlinks". En: *European Public Sphere: How much of it do we have and how much do we need?*. Mannheim: University of Mannheim. 213-264.

Leavitt, H. J. (1951). "Some effects of certain communication patterns on group performance". *The Journal of Abnormal and Social Psychology*, 46(1), 38.

Lee, J. (2006). "Computer-mediated communication as political communication: investigating the agenda-setting function". *Florida State University*.

Lippmann, W. ([1922] 2003). *La opinión pública*. San Lorenzo del Escorial: Langre.

Liu, B. (2007). *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media.

López García, G. (2006). "Comunicación en red y mutaciones de la esfera pública". *Zer*, 20, 231-249.

Lowi, T. J. (1967) "Party, Policy and Constitution in America". W.N. Chambers and W.D. Burnham (eds) *The American Party Systems: Stages of Development*. New York: Oxford University Press.

Luhmann, N. (1968). *Vertrauen: Ein Mechanismus der Reduktion sozialer Komplexität*, Stuttgart: Enke.

- Luhmann, N. (1974). "Öffentliche Meinung". Langenbucher, W. R. *Zur Theorie der politischen Kommunikation*. München: Piper, 27–54.
- Marletti, C. (1985). *Prima e dopo. Tematizzazione e comunicazione politica*. Turín: Eri.
- McBryan, O. A. (1994). GENVL and WWW: Tools for taming the web. *Proceedings of the first international world wide web conference*, 341.
http://www.iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel05/References/Pinkerton_1994/WebCrawler.html
- Nazar, R. (2011). A statistical approach to term extraction. *IJES, International Journal of English Studies*, 11(2), 159-182.
- Nazar, R., Vivaldi, J., & Wanner, L. (2012). Automatic taxonomy extraction for specialized domains using distributional semantics. *Terminology*, 18(2), 188-225.
- General Electric Company; McCall, J. A., Richards, P. K.; Walters, G. F. (1977). *Factors in Software Quality: Final Report*. Information Systems Programs, General Electric Company.
- Haken, H. (2006). *Information and self-organization: A macroscopic approach to complex systems*. Berlín: Springer-Verlag.
- McCombs, M.E.; Shaw, D.L. (1972). "The Agenda-Setting Function of Mass Media". *Public Opinion Quarterly*, 36(2), 176-187.
- Mills, C.W. (1958). "The structure of power in american society". *The british journal of sociology*, 9(1), 29-41. <http://www.csub.edu/~akebede/SOC502Mills2.pdf>
- Monge, P. R.; Contractor, N. S. (2003). *Theories of communication networks*. Oxford University Press.
- Ntoulas, A., Cho, J.; Olston, C. (2004). What's new on the web?: the evolution of the web from a search engine perspective. *Proceedings of the 13th international conference on World Wide Web*, 1-12. ACM.
- Olston, C.; Najork, M. (2010). Web crawling. *Foundations and Trends in Information Retrieval*, 4(3), 175-246. <http://msr-waypoint.com/pubs/121136/1500000017.pdf>
- Ortega, M., Pérez, M., & Rojas, T. (2003). Construction of a systemic quality model for evaluating a software product. *Software Quality Journal*, 11(3), 219-242.
http://lisi.usb.ve/publicaciones/02%20calidad%20sistemica/calidad_10.pdf
- Pant, G.; Srinivasan, P. (2005). Learning to crawl: Comparing classification schemes. *ACM Transactions on Information Systems (TOIS)*, 23(4), 430-462.
<http://home.business.utah.edu/actqp/Papers/Learning%20to%20Crawl%20Comparing%20Classification%20Schemes.pdf>
- Park, H.W. (2003). "Hyperlink Network Analysis: A New Method for the Study of Social Structure on the Web." *Connections*, 25(1), 49-61.

Park, H.W.; Thelwall, M. (2008). "Link analysis: Hyperlink patterns and social structure on politicians web sites in South Korea". *Quality and Quantity*, 42(5), 687-697.

Pinkerton, B. (1994, October). Finding what people want: Experiences with the WebCrawler. In *Proceedings of the Second International World Wide Web Conference*, 94, 17-20.

Pont-Sorribes, C. (2012). Las nuevas tecnologías y la percepción de los ciudadanos ante la comunicación de emergencias. *Anuario Hipertext.net*; 10.

Pont-Sorribes C, Cortiñas-Rovira S. (2013) New Media, Old Journalistic Work Routines. *Journal of Mass Communication & Journalism*; 0-0.

Pont-Sorribes, C; Codina, L; Pedraza-Jimenez, R. (2009) Comunicación de riesgo y sistemas de información en la Web: cinco modelos. *El Profesional de la información*; 18(4): 389-397.

Rhodes, R.A.W.; Marsh. D. (1992). "Policy Network in British Politics". Marsh and Rhodes (eds.). 1992a. 1-26.

Ripley, R. B.; Franklin, G. (1975). *Policy Making in the federal executive branch*; Ed. The Free Press, USA

Rodrigo, M. (2001). *Teorías de la comunicación: ámbitos, métodos y perspectivas*. Bellaterra: Universitat Autònoma de Barcelona. Servei de Publicacions.

Rodríguez-Martínez, R.; Pedraza-Jiménez, R. (2009). "Prensa digital y Web 2.0" [en línea]. "Hipertext.net", 7. <http://www.upf.edu/hipertextnet/numero-7/prensa-digital.html>

Rogers, R. (2002). "Operating Issue Networks on the Web". *Science as Culture*, 11(2), pp. 191-213.

Rogers, R.; Marres, N. (2000). "Landscaping climate change: a mapping technique for understanding science and technology debates on the world wide web". *Public Understanding Science*, 9(2), 141-163.

Rogers, R.; Ben-David, A. (2008). "The Palestinian-Israeli peace process and transnational issue networks: the complicated place of the Israeli NGO". *New Media & Society*, 10(3), 497-528.

Saperas, E. (1987). *Los efectos cognitivos de la comunicación*. Barcelona: Ariel.

Sampedro, V. (2000). Opinión pública y democracia deliberativa. Medios, sondeos y urnas. *International Journal of Media and Cultural Politics*, 3(1).

Sampedro, V. y Resina de la Fuente, J. (2010). Opinión pública y democracia deliberativa en la Sociedad Red1 *Ayer* 80(4), 139-162.

Schlozman, K. L.; Tierney, J. T. (1986). *Organized interests and American democracy*. Nueva York: Harper & Row.

Shaw, D. L.; Stevenson, R. L.; Hamm, B. J. (2002). "Agenda Setting Theory and Public Opinion Studies in a Post-Mass Media Age". *Egyptian Journal of Public Opinion Research*, 3(1), 1-20.

Shumate, M. (2012) "The Evolution of the HIV/AIDS NGO Hyperlink Network." *Journal of Computer-Mediated Communication*, 17(2), 20-134.

Shumate, M.; Dewitt, L. (2008). "The North/South Divide in NGO Hyperlink Networks." *Journal of Computer-Mediated Communication*, 13(2), 405–428.

Shumate, M.; Lipp, J. (2008). "Connective collective action online: An examination of the hyperlink network structure of an NGO issue network". *Journal of Computer-Mediated Communication*, 14(1), 178-201.

Schmidt-Manz, N.; Gaul, W. (2005). "Web mining and online visibility". *Classification, the ubiquitous challenge: 28th Annual Conference of the Gesellschaft. Klassifikation*. University of Dortmund.

SocSciBot [software] (2003). Statistical Cybermetrics Research Group
<http://socscibot.wlv.ac.uk/>

Solé, R. (2009). *Redes complejas: del genoma a Internet*. Barcelona: Tusquets.

Strogatz, S.H. (2001). "Exploring complex networks". *Nature*, 410(6825), 268-276.

Thelwall (2004a). Webometric Analyst [software]. Disponible en
www.webometrics.wlv.ac.uk

Thelwall (2004b). SocSciBot [software]. Disponible en <http://socscibot.wlv.ac.uk>

Thelwall, M. (2006). "Interpreting social science link analysis research: A theoretical framework". *Journal of the American Society for Information Science and Technology*, 57(1), 60-68.

Thelwall, M. (2012). A history of webometrics. *Bulletin of the American Society for Information Science and Technology*, 38(6), 18-23. https://asis.org/Bulletin/Aug-12/AugSep12_Thelwall.pdf

Thelwall, M.; Sud, P. (2011). A comparison of methods for collecting web citation data for academic organizations. *Journal of the American Society for Information Science and Technology*, 62(8), 1488-1497.

Thelwall, M.; Sud, P. (2012). Webometric research with the Bing Search API 2.0. *Journal of Informetrics*, 6(1), 44-52.

- Tordera, M. P. (2009). La complejidad en el asma: inflamación y redes libres de escala. *Archivos de Bronconeumología*, 45(9), 459-465.
- Vallez, M. et al. (2010). "Procedures for extracting keywords from web pages, based on search engine optimization". *Hipertext.net*, 8, http://www.upf.edu/hipertextnet/en/numero-8/keywords_extraction.html
- Vanderose, B. (2012). *Supporting a model-driven and iterative quality assessment methodology: The MoCQA framework*. FUNDP.
- Vanderose, B.; Ayed, H.; Habra, N. (2014). "Software Quality in an Increasingly Agile World". *ERCIM News*, (99), 36-37.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*, 8. Cambridge University Press.
- Watts, D. (2003). *Six degrees: The science of a connected age*. New York: Norton.
- Watts, D.; Strogatz, S.H. (1998). "Collective dynamics of 'small-world' networks". *Nature*. 393(6684), 440-442.
- Watts, D. J., & Dodds, P. S. (2007). Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4), 441-458.
- Wilkinson, K., Lowe, P., Donaldson, A. (2010). Beyond policy networks: policy framing and the politics of expertise in the 2001 foot and mouth disease crisis. *Public administration*, 88(2), 331-345.
- Young, L.E.; Leonardi, P.M. (2012). "Social Issue Emergence on the Web: A Dual Structural Model". *Journal of Computer-Mediated Communication*, 17(2), 231-246.

