



Análisis de redes : introducción

Xavier Polanco

► To cite this version:

Xavier Polanco. Análisis de redes : introducción. Mario Albornoz; Claudio Alfaraz. Redes de conocimiento: Construcción, dinámica y gestión, RICYT/CYTED/UNESCO Edición, pp.77-112, 2006, 978-987-98831-1-2. <hal-00218397>

HAL Id: hal-00218397

<https://hal.science/hal-00218397v1>

Submitted on 27 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Análisis de redes: introducción¹

Xavier Polanco

1. Introducción

2. Contexto

- 2.1. Redes
- 2.2. Estructura, contenido, uso
- 2.3. Dinámica y evolución

3. Una red de “ley de potencia”.

- 3.1 Características de las distribuciones de ley de potencia
- 3.2 Leyes y distribuciones hiperbólicas
- 3.3 Sistemas no lineales y altamente distribuidos

4. Modo de desarrollo

- 4.1 Efecto de red
- 4.2 Enlazado preferencial
- 4.3 Atractivo
- 4.4 Desviaciones

5. Elementos de la teoría de grafos

- 5.1 Definición de grafo
- 5.2 Tipos de grafos
- 5.3 Conceptos
- 5.4 Tipos de relaciones
- 5.5 Tipos de nodos

6. Pequeño mundo

- 6.1 La aparición del pequeño mundo
- 6.2 Propiedades y definición
- 6.3 El pequeño mundo en la Web

7. Centralidad, centralización y densidad

- 7.1 Grado, intermediación, proximidad
- 7.2 Centralidad
 - 7.2.1 Centralidad de grado
 - 7.2.2 Centralidad de intermediación
 - 7.2.3 Centralidad de proximidad
- 7.3 Centralización
- 7.4 Densidad

8. Modelos de análisis de redes de conocimiento

- 8.1 Redes de conocimientos
- 8.2 Modelo de la citación
 - 8.2.1 Centros de actividad (*hubs*) y autoridades
 - 8.2.2 Web comunidades
 - 8.2.3 PageRank
- 8.3 Modelo de la co-ocurrencia o asociación
 - 8.3.1 Método general
 - 8.3.2 Análisis de co-sitios
 - 8.3.3 Sobre páginas Web
 - 8.3.4 Utilización del conocimiento
- 8.4 Modelo de los mapas de auto organización
 - 8.4.1 SOM
 - 8.4.2 Multi-SOM
 - 8.4.3 Aplicaciones

9. Conclusión

Referencias

¹ Mario Albornoz y Claudio Alfaraz, editores, *REDES DE CONOCIMIENTO: CONSTRUCCIÓN, DINÁMICA Y GESTIÓN*. Primera edición: Agosto de 2006 - ISBN-10: 987-98831-1-X ISBN-13: 978-987-98831-1-2. Pp. 77-112
Correo electrónico: ricyt@ricyt.org Sitio web: <http://www.ricyt.org>

1. Introducción

A partir del hecho que las redes existen y que ellas son de diferentes naturaleza, aquí nos interesamos por tres clases de redes: la red tecnológica de la comunicación, representada por Internet y la Web, la red social íntimamente ligada a la primera, y formada por el conjunto de usuarios de estas tecnologías en sus intercambios de información y de conocimientos, los cuales se configuran a su vez como redes de conocimiento. Digamos que la llamada “sociedad de la información” resulta por una gran parte del encuentro de estas tres redes en una estructura global.

Que hablemos de “sociedad de la información” o de “sociedad del conocimiento”, privilegiando en un caso la noción de información y en el otro la de conocimiento, bien que nociones no necesariamente simétricas, ellas comportan la propiedad común de constituir una sociedad que podemos llamar “sociedad en redes” (la *network society* de Castells, 1996), subrayando así que el principio de red (*network*) está configurando de una manera dominante, como nunca antes, la estructura de la sociedad contemporánea. Así, por ejemplo, el número de ocurrencias de la palabra “red” en la literatura del “management” se multiplica por más de 20 entre los años 1960 y 1990 (Boltanski & Chiapello, 1999). Por otra parte, y desde siempre, las estructuras tecnológicas (digamos caminos, ferrocarriles, transporte aéreo, electricidad, telecomunicaciones,...) han sido analizadas en ingeniería utilizando entre otras la teoría de grafos. Tanto la información como el conocimiento pueden ser descritos en términos de redes. Sobre todo si admitimos que la búsqueda como la emisión de informaciones o de conocimientos se realiza hoy en día en el marco de una situación social significada por *CMC*, es decir, “*computer mediated communication*” (Garton et al, 1997).

Como sabemos, las redes sociales son el objeto de estudio de lo que se conoce en ciencias sociales como el “análisis de redes sociales” (Knoke & Kuklinski, 1982, Scott, 1991, Wasserman & Faust, 1991, Degenne & Forsé, 2001)². Pero es interesante de observar que el “análisis de redes sociales” ha sido redescubierto y retomado por los analistas de la Web considerándola como un caso de red social. En efecto, existe un cierto consenso en cuanto a la idea que la Web puede ser considerada desde el paradigma del análisis de las redes sociales, esto es, siguiendo el enfoque desarrollado en el campo de las ciencias sociales (Mercklé, 2004), y que es así extendido al análisis de las redes tecnológicas de la comunicación mediante las cuales se constituyen las redes de conocimiento.

La noción general de red que aquí utilizaremos hace referencia al menos a dos conjuntos de información: un conjunto de elementos representados como nodos, y un conjunto de relaciones que pueden ser orientadas o no; también es posible que haya un conjunto de valores numéricos o intencionales caracterizando las relaciones. Esta noción general puede formularse matemáticamente mediante la teoría de grafos. Un espectro amplio de sistemas en física, biología, sociedad y tecnología son descritos como redes complejas. Los ejemplos van de las proteínas y la biología celular a las tecnologías de la comunicación tales que Internet y la Web.

Además de su modelización en términos de grafos, muchas de estas redes complejas aparecen gobernadas por ciertos principios de organización estudiados por la física estadística. Y por lo que aquí nos interesa, éstos ya han sido aplicados al estudio de la estructura y dinámica de la

² Consultar: <http://www.insna.org/>

Web (Albert & Barabási, 2000; Barabási, 2001; Huberman, 2001). El modelo es aquí el de un sistema no lineal que obedece a una “ley de potencia libre de escala”.

Nuestro objetivo es entonces de mostrar que la red tecnológica representada por Internet y la Web, obedece al modelo de los sistemas no lineales y de ley de potencia (sección 3), y que presenta un modo de desarrollo característico (sección 4). Siguiendo el ejemplo del análisis de redes sociales, la Web puede ser analizada, de acuerdo a la teoría de grafos, como un grafo pequeño mundo (secciones 5 y 6), lo cual significa que podemos aplicar las nociones de centralidad y de densidad (sección 7). Además, los conocimientos se organizan en la forma de redes que pueden ser analizadas de acuerdo con los ejemplos paradigmáticos de los análisis de citas y de las palabras asociadas (sección 8). En suma, el objetivo es de proponer, de acuerdo con el estado del arte, modos de analizar las redes, tecnológica de la Web, social de sus usuarios y de los conocimientos contenidos en la Web, asimismo que las consecuencias que se derivan de las propiedades observadas. Me referiré en seguida al contexto en el cual nos situamos.

2. Contexto

Sea una red global, R_G , a la cual se refieren las expresiones “sociedad de la información” o “sociedad del conocimiento”. Este es nuestro contexto general. Citemos un documento de la Comisión europea describiendo las propiedades socio-económicas de la R_G en los términos siguientes:

“The fast development of the Information and Communication Technology (ICT) has brought about deep *changes in our way of working and living*, as the widespread diffusion of ICT is accompanied by organizational, commercial, social and legal innovations.

Our society is now defined as the “Information Society”, a society in which low-cost information and ICT are in general use, or as the “Knowledge(-based) Society”, to stress the fact that the most valuable asset is investment in intangible, human and social capital and that the key factors are knowledge and creativity”.³

Esta citación me parece que ella expresa bien las características socio-económicas de la red global R_G ; además ella propone distinguir información y conocimiento. Por otra parte, Garton et al (1997) observan que una red informática constituye una red social cuando ella conecta personas u organizaciones. Además, estos autores hacen uso de un concepto que me parece importante y es el de “computer-mediated communication” (*CMC*), para significar la imbricación de una red tecnológica (como la Web) con la red social de los usuarios.

2.1. Redes

Ahora bien, esta red global R_G está formada por un conjunto de redes específicas, y teniendo en cuenta el contexto socio-técnico en el cual nos colocamos, *CMC*, podemos limitar a las siguientes: 1) la red tecnológica de la comunicación, R_{TC} , es decir, la Web e Internet, 2) la red social, R_S , de los actores humanos de la red tecnológica, 3) la red económica, R_E , generada por el uso de la R_{TC} en la esfera de la actividad económica, 4) la red de los conocimientos, R_C , producidos por los usuarios de la R_{TC} y que consideraremos como el contenido de ella.

³ Consultado (15/03/06): http://europa.eu.int/comm/employment_social/knowledge_society/index_en.htm

La formula es entonces: $[R_G (R_{TC}, R_S, R_E, R_C)]$, siendo R_{TC} igual a Web/Internet, es decir, la Web es la superestructura e Internet la infraestructura de R_{TC} .

Como he dicho en la introducción, los estudios de la Web convergen en tratarla de acuerdo con el análisis de redes sociales (Chakrabarti, 2000, Kumar, 2002). Y como sabemos (Wasserman & Faust, 1999, capítulo 4), la teoría de grafos es ampliamente utilizada por el análisis de redes sociales para representar formalmente las relaciones sociales y cuantificar importantes propiedades de las estructuras sociales, así como también ahora en lo que se refiere a la estructura de la Web (Chakrabarti et al., 1999; Broder et al., 2000; Barabási, 2001).

Además de su utilidad como sistema matemático, la teoría de grafos permite la representación de una red social como un modelo que consiste en un conjunto de actores y los lazos entre ellos. Cuando un grafo es utilizado como modelo de una red social, los nodos representan los actores, y las relaciones conectando los nodos significan los lazos entre los actores. Así, el modelo consiste en un conjunto de nodos y un conjunto de relaciones. Al igual que las redes sociales, la Web, red tecnológica, puede ser representada como un conjunto de nodos y un conjunto de lazos hipertextos (relaciones) entre los nodos (paginas, sitios).

A propósito de R_C , es decir la red de conocimientos, encontramos las relaciones que significan las citaciones creando, como ya lo había observado Price (1965), una red de publicaciones científicas, de suerte que el análisis de citaciones sería entonces el medio de hacer evidente el sistema de relaciones constitutivas de esta red. El objetivo de su artículo de 1965, “*Network of Scientific Papers*”, es como él mismo dice: “describir la naturaleza de la red mundial de artículos científicos en sus líneas generales”. Más tarde se hablará de esta red en términos de indicadores relacionales, estimándose que las citaciones y las co-citaciones son indicadores relacionales de primera generación, y las palabras asociadas indicadores relacionales de segunda generación (ver a este propósito Callon et al, 1993). Así, el objetivo de los análisis de citaciones, co-citaciones y palabras asociadas es el de poner en evidencia estas redes (sistemas de relaciones). A ello viene agregarse el hecho que con la Web e Internet la R_C se encuentra en la trama misma de la R_{TC} .

2.2. Estructura, contenido, uso

La Web en tanto que red posee una estructura en la cual se encuentra un contenido como también determina un uso, una utilización que ella registra como información. De suerte que su análisis puede concentrarse sea sobre la estructura, sea sobre el contenido, sea sobre el uso que de ella se hace. Además, la red que la Web representa se caracteriza por una dinámica, un tipo de evolución o modo de desarrollo.

El análisis de la Web puede entonces declinarse según tres categorías de análisis: de la estructura, del contenido, del uso (como dicen Kosala & Blockeel, 2000, en el campo del “*Web Mining*”). De acuerdo con estos autores, el análisis de la estructura de la Web trata de descubrir el modelo subyacente a las estructuras de hiperlazos. El modelo se basa en la topología de los hiperlazos. El análisis del contenido de la Web busca descubrir la información útil (a una necesidad). El contenido de la Web consiste en varios tipos de datos, textual, imagen, audio, video, meta-datos. Mucho de los datos contenidos en la Web, y son los que más nos interesan aquí, son datos textuales no estructurados. El análisis del uso de la Web

trata los datos generados por el comportamiento de los usuarios en las sesiones y que son registrados en los ficheros “log” de los servidores.

Por su parte Ingwersen & Björneborn (2004) señalan cuatro campos de análisis en el estudio infométrico de la Web y que recubren los que hemos visto en “*Web Mining*”, es decir, el análisis de contenido de las páginas Web, el análisis de la estructura de enlaces, el análisis del uso de la Web, como también el análisis de la tecnología de la Web; esta visión expresa el punto de vista de aquellos que desde la “library and information science” analizan en términos cuantitativos la Web bajo el nombre de “*Webometrics*”.

2.3. Dinámica y evolución

La red tecnológica de la Web, constituye una estructura evolutiva, un sistema dinámico, lo cual significa que además del análisis de la estructura es necesario tener en cuenta su dinámica de evolución o modalidad de desarrollo. Esta dinámica presenta propiedades sumamente interesantes y que vamos abordar a continuación (en las secciones 3 y 4). Se trata de un sistema dinámico no lineal obedeciendo una ley de potencia, lo cual nos lleva a observar el parentesco con la naturaleza de las leyes bibliométricas, es decir, las distribuciones hiperbólicas conocidas desde hace ya largo tiempo en el campo de la información bibliográfica.

3. Una red libre de escala y ley de potencia.

Las distribuciones observadas en la Web obedecen a una “ley de potencia” (*power-law*). Su definición es $1/n^\beta = n^{-\beta}$. Lo que quiere decir que la probabilidad de encontrar un sitio Web con un número de enlaces y/o páginas, n , es igual a $1/n^\beta$ siendo $\beta \geq 1$. Se trata entonces de distribuciones hiperbólicas representadas por curvas que presentan las formas que vemos en la figura 1. La distribución “libre de escala y por ley de potencia” aparece como una propiedad común a muchas grandes redes.

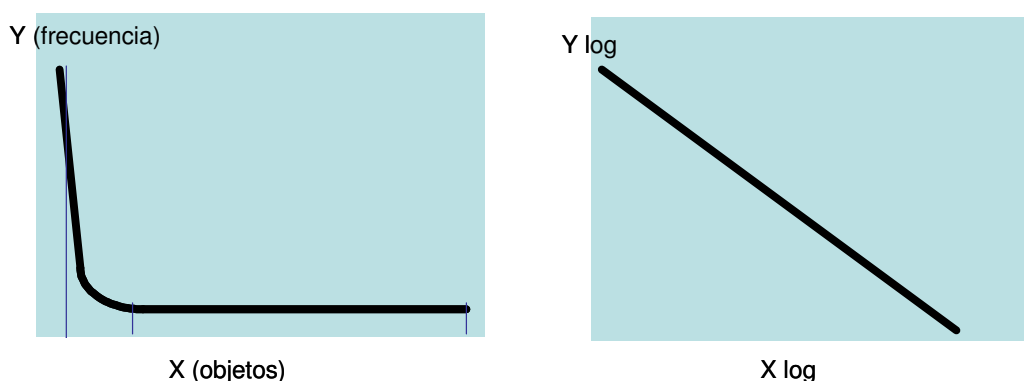


Figura 1: Curvas representando una distribución hiperbólica o por ley de potencia. Lo mismo puede expresarse por una curva de Lorenz y entonces podemos aplicar el índice de concentración de Gini, particularmente bien adaptado para el estudio de las distribuciones fuertemente asimétricas en donde la relación a la media no tiene gran sentido. El índice de Gini, comprendido entre 0 y 1, tiene un valor tanto más elevado que la distribución es desigual.

Los estudios estadísticos de la Web muestran que a pesar de la aparente arbitrariedad de su desarrollo, la red obedece a patrones (*patterns*) que reflejan la existencia de regularidades no manifiestas (*hidden regularities*), como la distribución por ley de potencia, es decir, la existencia de muchos pequeños sitios en la Web y de pocos grandes en número de páginas o

de enlaces, pocas páginas contienen millones de enlaces, pero una gran cantidad de páginas tienen uno o dos enlaces. Esta distribución se expresa matemáticamente, como hemos dicho, en términos de la ley de potencia, es decir, $1/n^\beta$. La distribución del número de páginas por sitio, y del número de enlaces (“*links*”) saliendo de un sitio o viniendo a este mismo sitio, obedece a la distribución por ley de potencia. Los estudios estadísticos de la estructura característica de la Web han puesto en evidencia esta regularidad (como lo resume Huberman, 2001, en el capítulo 3 de su libro sobre las leyes de la Web). Más adelante consideraremos las desviaciones observadas con respecto al modelo de la ley de potencia (sección 4.4).

3.1 Características

Las distribuciones libre de escala y por ley de potencia presentan al menos estas tres propiedades: una forma muy dispareja de distribución, representadas por curvas de pendiente muy pronunciada y colas muy largas, y en las cuales el comportamiento medio no es significativo.

a) Una red libre de escala (*scale-free network*) es una red compuesta de nodos y enlaces, que tienen la particularidad de que los enlaces están distribuidos de forma muy dispareja, se llama “libre de escala”, porque en estos tipos de redes, a menudo se observa, que un nodo crece (en términos de enlaces) proporcionalmente al tamaño que tiene, sin que haya un parámetro de escala que indique, por ejemplo, que dado un cierto número de enlaces ya no se pueden ganar más enlaces o deben agregar se más lentamente.

b) Las curvas de las distribuciones por ley de potencia presentan colas muy largas, lo cual significa que, en la Web, hay una probabilidad limitada de encontrar sitios que sean extremadamente grandes comparados a la media de los sitios. Es decir, la distribución no es normal o gaussiana como es habitual de encontrar en el mundo. Una distribución por ley de potencia, a diferencia de la forma en campana de las distribuciones normales, no es simétrica en torno a su máximo, presentando una fuerte pendiente y concluyendo en una larga cola (“*skewed*”, Simon, 1957).

El comportamiento medio del sistema no es significativo, la noción de media no tiene sentido, no es típica. Un tamaño típico o tipo es aquel que encontramos más frecuentemente en una muestra. No existe algo típico en la red libre de escala. Si bien podemos sacar un promedio, el promedio no sirve para nada porque estas redes tienen elementos con muchísimas relaciones y elementos con muy pocas relaciones, sin que exista una escala característica de la red completa. Así, por ejemplo, lo que vamos encontrar en los análisis del número de páginas o de enlaces en la Web, será que la mayoría de sitios contienen un número pequeño de páginas o enlaces que la esperada media.

3.2 Leyes y distribuciones hiperbólicas

El análisis de las propiedades estadísticas de la información científica ha formulado las conocidas leyes “bibliométricas” concerniendo las revistas (Bradford, 1934), los autores (Lotka, 1926), como también en lo que se refiere al lenguaje (Zipf, 1949); en la esfera de la economía tenemos la ley de Pareto (1897). Todas ellas tienen la propiedad común de ser hiperbólicas, y que Simon (1957) describió como “*a skew or hyperbolic distribution*” (a la cual se refiere por Price en su artículo de 1976 acerca de “una teoría general de la bibliometría y otros procesos de ventajas acumulativas”). Autores como por ejemplo Haitun (1982) y Yablonski (1985) han llegado a la conclusión de la existencia de un universo estadístico conforme al principio de “*non-Gaussian distribution*”. Los análisis de Internet y de la Web han demostrado que esta red tecnológica obedece a este mismo

universo estadístico, es decir, al principio de distribución libre de escala y por ley de potencia, que aparece como una propiedad común a muchas grandes redes. En efecto como ha podido comprobarse, no solo Internet (Faloutsos et al, 1999) sino que también la Web son ejemplos de redes libre de escala, es decir que ambos obedecen al principio de la distribución por ley de potencia (Huberman et al, 1998, Barabási & Albert, 1999, Albert et al, 1999; Huberman & Adamic, 1999, Huberman, 2001, Adamic & Huberman 2002).

3.3 Sistemas no lineales y altamente distribuidos

Además, la mayoría de las redes complejas del tipo de la Web e Internet constituyen sistemas no lineales y altamente distribuidos. Lo cual quiere decir que no podemos establecer en el análisis una conexión entre la acción de los actores individuales y las formas de comportamiento del sistema en su globalidad. En otras palabras, el comportamiento de la red no puede ser explicado por medio de la adición o suma de las acciones parciales de sus constituyentes. Esta situación define lo que se llama un sistema no lineal. En suma, una red del tipo de la Web implica estas dos propiedades: ser no lineal y al mismo tiempo altamente distribuido. Por eso un conocimiento detallado de como un determinado sitio evoluciona por la adición de nuevas páginas y enlaces, no es suficiente como para comprender el desarrollo y la estructura de la red en su conjunto.

Los sistemas no lineales se caracterizan por comportamientos aleatorios aun cuando su descripción matemática sea determinista. Esta situación se llama caótica y en tanto que tal ella es el objeto de estudio de la llamada “teoría del caos” (véase Ruelle, 2000, para una introducción no técnica). Es decir, si uno considera el estado del sistema en su punto de partida de acuerdo con determinadas condiciones iniciales, y sigue su evolución en el tiempo, el resultado final es bastante diferente a lo previsto. La gran sensibilidad a las condiciones iniciales aparece ser extremadamente aleatoria, y las solas predicciones que pueden ser hechas a propósito de sus comportamientos son probabilísticas.

4. Modo de desarrollo

El análisis estructural de una red no es suficiente, es necesario además considerar su dinámica, es decir, sus cambios durante el tiempo en la estructura de los enlaces. Los modelos libres de escala consideran las redes como sistemas dinámicos que se auto-organizan y evolucionan en el tiempo por medio de la adición y desaparición de nodos y enlaces.

El modo de desarrollo de una red como la Web obedecería a tres factores principales: 1) el número de sitios, páginas y enlaces, sigue una distribución por ley de potencia, 2) los sitios aparecen a momentos diferentes en el tiempo, y por lo tanto deben considerarse los diferentes momentos de partida y además, 3) ciertos sitios crecen más rápido que otros.

El factor 1 supone la existencia de un mecanismo universal que sostenga el desarrollo de la Web y produzca la distribución por ley de potencia. Por ejemplo, el número de páginas o de enlaces de un sitio, n , en un instante dado, es igual al número de páginas o enlaces del sitio en el instante anterior, más o menos una fracción probable de n . En otras palabras, el tamaño del sitio determina la dimensión del desarrollo esperado. En realidad, este factor solo conduce a una distribución log-normal que no es una distribución por ley de potencia (hiperbólica). Es necesario, para que haya una distribución por ley de potencia, considerar los otros dos factores (2 y 3) que influyen en el desarrollo de la Web.

En conclusión, el desarrollo aleatorio multiplicativo (factor 1), combinado con el hecho que los sitios aparecen a diferentes momentos en el tiempo (factor 2) y que crecen a tasas diferentes (factor 3), conduce a una explicación del comportamiento dominante en la Web en términos del modelo de la ley de potencia libre de escala.

La regularidad de la ley de potencia libre de escala concierne la distribución de páginas y de enlaces por sitio. Este modelo explica bien matemáticamente los datos empíricos acerca del desarrollo de la red; en cada periodo de tiempo el número de nuevos enlaces que un sitio recibe es una fracción aleatoria del número de enlaces que el sitio ya posee en el momento considerado. En otras palabras, la tasa de adquisición de nuevos enlaces es proporcional al número de enlaces que el sitio ya posee. Otro punto interesante es que no habría correlación entre la edad del sitio y su número de enlaces. Además, la tasa de crecimiento varía de sitio en sitio.

Estaríamos delante de una red cuyo modo global de desarrollo es del tipo Zipf-Pareto, y que es interesante de comparar con el proceso que Price (1976) ha llamado *CAD*, es decir, “*Cumulative Advantage Distribution*”, y propuesto como un modelo estadístico de la situación en la cual “*success breeds success*”. Por su parte, Mandelbrot (1954, 1967, 1968, 1975) ha formalizado dicho tipo de distribución tanto en el dominio de las ciencias sociales que de la lingüística estadística, dando origen a la llamada “ley Mandelbrot-Zipf”.

4.1 Efecto de red

Un punto importante a tener en cuenta es el “efecto de red” formulado por la “ley de Metcalfe”,^{4,5} ella establece que el valor de una red crece en proporción al cuadrado del número de usuarios, y cuya fórmula sería $U(U-1) = (U^2 - U)$, siendo U el nombre de usuarios. La red crece en nodos y enlaces y así su valor aumenta para cada usuario, pero el valor total de la red aumenta mucho más rápido que el número de usuarios. Esto es conocido como el efecto de red.⁶ Desde un punto de vista económico, el efecto de red contrasta con los modelos tradicionales de la ley de la oferta y la demanda.

4.2 Enlazado preferencial

El mecanismo llamado enlazado preferencial (*preferential attachment*), o conexión preferencial, designa el proceso por el cual “*rich-get-richer*” (Barabási & Albert, 1999, Huberman & Adamic, 1999, Kumar et al., 2000, Albert & Barabási, 2002), es decir, la red crece por la llegada de nuevos nodos, y estos nodos enlazan con preferencia a aquéllos que ya tienen enlaces. Así, el mecanismo del enlazado preferencial significa que la probabilidad que un nodo existente gane enlaces es proporcional al número de enlaces que este nodo ya tiene en ese momento. En el modelo libre de escala, como explica Barabási (2001), la tasa que un nodo adquiera nuevos lazos está dada por $dk / dt = mP_i(k)$, en donde m es el número de lazos que un nuevo nodo posee cuando se integra a la red, esta fórmula pronostica que cada nodo aumenta su conectividad en el tiempo de acuerdo a la ley de potencia: $k(t) = t^\beta$, en donde $\beta = 1/2$ es el exponente dinámico.

⁴ Véase http://en.wikipedia.org/wiki/Metcalfe's_law

⁵ Véase <http://americanhistory.si.edu/collections/comphist/montic/metcalfe.htm#me7>

⁶ Véase http://en.wikipedia.org/wiki/Network_effect

4.3 Atractivo

En el análisis de la Web, se ha comprobado que la habilidad de los nodos a competir por enlaces, a expensas de otros nodos, varía en función de un índice de atractivo (*fitness*), η , para todo nodo i , de suerte que la probabilidad, $P_i(k_i)$, que un nuevo nodo se conecte a uno con k_i enlaces, se modifica de acuerdo con $P_i(k_i) = \eta_i k_i / \sum \eta_j k_j$. Un tal índice de atractivo supone, empíricamente, que un sitio (nodo) presente las propiedades siguientes: buen contenido, buena actualización del contenido, una interface amigable, y marketing. La reunión de estas propiedades contribuye a que su índice de atractivo sea mayor.

La competición generada por los diferentes niveles de atractivo significa que cada nodo evoluciona diferentemente en el tiempo comparado a otros. La conectividad de cada nodo es dada por $k_i(t) \equiv t^{\beta(\eta)}$ (Albert & Barabási 2002), en donde el exponente $\beta(\eta)$ aumenta con η . Los nodos atractivos, aquéllos con elevado η , pueden integrarse a la red algún tiempo más tarde y conectarse con muchos más enlaces que los nodos menos atractivos y que forman parte de la red después de mucho más tiempo.

4.4 Desviaciones

Estudios recientes buscan mejorar la precisión del modelo original del enlazamiento preferencial (Pennock et al., 2002; Albert & Barabási, 2000). Puesto que hay situaciones reales en las cuales el comportamiento previsto por el modelo de la ley de potencia libre escala no se realiza. Así, por ejemplo, se ha encontrado que la probabilidad, $P_i(k)$, que un nuevo nodo se conecte a uno con k enlaces puede ser no lineal, lo que conduce a una desviación de lo previsto por la ley de potencia (Krapivsky et al., 2000). Además, desviaciones ocurren cuando se consideran pequeños números de enlaces (Broder et al., 2000); y las desviaciones varían según las diferentes categorías de páginas (Pennock et al., 2002), por ejemplo, la distribución de enlaces de los sitios de universidades diverge fuertemente de la ley de potencia, siguiendo una distribución mucho más uniforme, y como Pennock et al (2002) notan, en este sector “*winner don’t take all*”.

Enlaces se agregan a los nodos existentes de la red, o nodos y enlaces pueden desaparecer, Dorogovtsev et al. (2000) han demostrado que la presencia de tales acontecimientos puede modificar el exponente de la ley de potencia, conduciendo en la práctica a cualquier valor entre uno e infinito. Amaral et al. (2000) han puesto en evidencia que el envejecimiento y los efectos de saturación limitan los enlaces que un nodo puede adquirir, induciendo un corte exponencial en $P(k)$. Barabási (2001) observa que la regularidad de la ley de potencia se ajusta bien con fenómenos físicos críticos como por ejemplo la congelación del agua, o el ordenamiento de los spins en un magneto, pero que existe una diferencia crucial entre esos fenómenos y los sistemas evolutivos. En los fenómenos críticos los exponentes son estables y universales, y no pueden ser cambiados por la modificación de ciertos parámetros. En cambio, en la Web, el exponente puede variar continuamente modificando casi todos los parámetros que gobiernan los enlaces y nodos. Así la universalidad conocida en física estaría aquí ausente.

NOTA: Me parece importante observar, al paso, que las redes obedecerían, de acuerdo a los que hemos visto, a un mecanismo significado por las expresiones lapidarias: “success-bread-success”, “rich-get-richer”, o “winner-take-all”, y que por consecuencia las distribuciones y el desarrollo de las redes obedecerían a esta sola modalidad. En la sociología de las ciencias, Merton (1973) expone el llamado “efecto San Mateo” (*Mathew Effect*), y Price (1976)

subraya la existencia de una economía de ventajas cumulativas en el campo de la ciencia, basándose justamente en las leyes bibliométricas (Bradford, Lotka, Zipf) como también en la ley de Pareto. Además, se habla en sociología general del modelo de “*winner-take-all society*” (Frank & Cook 1995), y que tendríamos que relacionar con la “*network society*” (Castells, 1996). Es cierto que existen en este contexto formulas más matizadas como “winner-take-almost” o “winner-don’t-take-all”. Dejo aquí sólo planteada esta cuestión. En efecto, podemos interrogarnos sobre las consecuencias sociales y políticas del predominio de las redes en la estructura de la sociedad contemporánea.

5. Elementos de la teoría de grafos

En esta sección, vamos a examinar la proposición que la teoría de grafos y sus métodos pueden ser usados para analizar la estructura de una red como la Web: las páginas y los enlaces son vistos como un conjunto de nodos y un conjunto de conexiones de un gran grafo. Aquí se trata de definir solamente algunos elementos de la teoría de grafos y que hemos aplicado en el análisis de la Web (Polanco, 2002; y al cual hago referencia en la sección 8.3.2), a saber los distintos tipos de grafos, de nodos y de relaciones, como instrumentos para un análisis descriptivo (como se hace en el análisis de las redes sociales).

5.1 Definición de grafo

Un grafo $G(N, R)$ consiste en dos conjuntos de información: un conjunto de nodos, $N = \{n_1, n_2, \dots, n_N\}$, y un conjunto de relaciones, $R = \{r_1, r_2, \dots, r_R\}$ entre pares de nodos.

En un grafo cada relación es un par no ordenado de nodos $r_k = (n_i, n_j)$. Una relación es no ordenada cuando la relación entre los nodos n_i y n_j es idéntica a la relación entre los nodos n_j y n_i , es decir, $(n_i, n_j) = (n_j, n_i)$. Estas relaciones son llamadas aristas. Una arista es entonces un par no ordenado de nodos y registra simplemente la presencia de un enlace entre dos nodos. Sea R el conjunto de aristas, y N el número de nodos, puesto que una arista es un par no ordenado de nodos, hay $N(N - 1) / 2$ posibles aristas en R .

5.2 Tipos de grafos

Además del tipo no orientado, y que viene de ser definido, existen otros dos tipos: los grafos ponderados y los grafos orientados.

Los *grafos ponderados* son la representación adecuada de las relaciones ponderadas o comportando un valor numérico. Así un grafo ponderado es un grafo en el cual cada relación lleva un valor o peso. Un grafo ponderado consiste entonces en tres conjuntos de información: un conjunto de nodos, $N = \{n_1, n_2, \dots, n_N\}$, un conjunto de relaciones, $R = \{r_1, r_2, \dots, r_R\}$, y un conjunto de valores o pesos, $W = \{w_1, w_2, \dots, w_R\}$, asociados a las relaciones. La notación de un grafo ponderado es $G_W(N, R, W)$.

En los *grafos orientados* las direcciones de las relaciones son especificadas. Estas relaciones orientadas son llamadas arcos. La diferencia entre las aristas y los arcos es que los arcos son pares ordenados de nodos y este orden refleja la dirección del enlace.

Un grafo orientado, $G_D(N, R)$, consiste en dos conjuntos de información: un conjunto de nodos, $N = \{n_1, n_2, \dots, n_N\}$, y un conjunto de arcos, $R = \{r_1, r_2, \dots, r_R\}$ entre pares de nodos. Cada arco es un par ordenado de nodos $r_k = \langle n_i, n_j \rangle$. La dirección del arco $\langle n_i, n_j \rangle$ va de n_i (el

origen o emisor) a n_j (el término o receptor). Sea R el conjunto de arcos y N el número de nodos, puesto que cada arco es un par ordenado de nodos, hay $N(N - 1)$ posibles arcos en R .

Se dice que un nodo es incidente con un arco, si el nodo esta en el par ordenado de nodos definiendo el arco. Ambos nodos n_i y n_j son incidentes con el arco $r_k = \langle n_i, n_j \rangle$. Y la adyacencia de dos nodos en un grafo orientado supone considerar si un nodo es primero o segundo en el par ordenado de nodos definiendo el arco. Formalmente, el nodo n_i es adyacente al nodo n_j si $\langle n_i, n_j \rangle \in R$, y el nodo n_j es adyacente desde el nodo n_i si $\langle n_i, n_j \rangle \in R$. En los diagramas el arco $\langle n_i, n_j \rangle$ será representado por una flecha yendo de n_i a n_j .

5.3 Conceptos

Todos los conceptos a los cuales aquí se hace referencia son extremadamente útiles cuando se trata de analizar cualquiera red cuyo modelo sea un grafo (como el que vemos en la figura 4):

- El “grado” de un nodo es el número de nodos adyacentes a este nodo, el número de relaciones incidentes con él. En un grafo orientado un nodo puede ser “adyacente a” o “adyacente desde” otro nodo, dependiendo de la dirección del arco.
- El “grado saliente” (*out-degree*) es el número de arcos originándose de n_i , y el “grado entrante” (*in-degree*) es el número de arcos terminando en n_i . El grado saliente de un nodo, $d_{out}(n_i)$, es el número de nodos adyacente desde n_i . El grado saliente del nodo n_i igual al número de arcos de la forma $r_k = \langle n_i, n_j \rangle$, para toda $r_k \in R$, y para todo $n_j \in N$. El grado entrante es un nodo, $d_{in}(n_i)$, es el número de nodos que son adyacentes a n_i . El grado entrante de un nodo n_i es igual al número de arcos de la forma $r_k = \langle n_j, n_i \rangle$, por toda $r_k \in R$, y por todo $n_j \in N$.
- Un “camino” es una sucesión de nodos tal que de cada uno de sus nodos existe una arista (o relación o conexión) hacia el nodo sucesor. Se dice que un camino es simple si no se repite ninguno de sus nodos en él.
- La “longitud de un camino” es el número de aristas que usa dicho camino.
- La “distancia geodésica” entre dos nodos es el menor número de aristas de un recorrido entre ellos.
- El “diámetro” de un grafo es la mayor distancia entre dos nodos del mismo.

5.4 Tipos de relaciones

Consideremos los tipos de relaciones (arcos) que pueden existir, asimismo que los tipos de nodos, por su utilidad en el análisis de una red que responde al modelo de un grafo orientado. Digamos que el tipo de enlace entre pares de nodos puede ser nulo, asimétrico o mutuo:

- *Nulo* es cuando ningún arco existe, cuando ninguno de los arcos $\langle n_i, n_j \rangle$ ni $\langle n_j, n_i \rangle$ esta contenido en el conjunto de relaciones orientadas R .
- Un par *asimétrico* de nodos posee un arco entre los dos nodos yendo en una u otra dirección, pero no en las dos direcciones simultáneamente, esto es, sólo uno de los arcos $\langle n_i, n_j \rangle$ o $\langle n_j, n_i \rangle$ está contenido en el conjunto de arcos R . Lo que normalmente se representa $\langle n_i \rightarrow n_j \rangle$ o $\langle n_j \rightarrow n_i \rangle$.
- Los pares *mutuos* o *recíprocos* de nodos tienen dos arcos entre ellos, uno yendo en una dirección y el otro viniendo en la dirección opuesta. Ambos arcos $\langle n_i, n_j \rangle$ y $\langle n_j, n_i \rangle$ se encuentran en el conjunto de relaciones R . Lo cual se representa gráficamente por una doble flecha $\langle n_i \leftrightarrow n_j \rangle$.

5.5 Tipos de nodos

Una red compuesta por enlaces orientados, como la Web, puede caracterizarse además por nodos aislados, transmisores, receptores y portadores. Los nodos aislados no presentan relaciones de ningún tipo, en cambio los llamados transmisores son nodos con relaciones originándose de ellos solamente, al contrario los receptores son nodos con solo relaciones terminando en ellos, y en fin los llamados portadores son los nodos originando y recibiendo relaciones o enlaces. Formalmente:

- Aislado : si $d_{in}(n_i) = d_{out}(n_i) = 0$
- Transmisor: si $d_{in}(n_i) = 0$ y $d_{out}(n_i) > 0$
- Receptor: si $d_{in}(n_i) > 0$ y $d_{out}(n_i) = 0$
- Portador: si $d_{in}(n_i) > 0$ y $d_{out}(n_i) > 0$

En suma, y como se ha dicho, todos estos elementos pueden ser utilizados para el análisis del grafo orientado que representa una determinada red (como muestra la figura 4).

6. Pequeño mundo

En general las redes son representadas por grafos que presentan la propiedad llamada “pequeño mundo” (*small world*). Es este otro concepto importante en el análisis de redes de acuerdo con la teoría de grafos. Y una de las cuestiones más importante acerca de la Web, en tanto que grafo, es justamente la existencia de una estructura pequeño mundo como en el caso de las redes sociales, en donde el problema del “pequeño mundo” puede formularse de la manera siguiente: dado dos individuos escogidos al azar en una población, cual es la probabilidad para que el número mínimo de intermediarios necesarios para conectarlos sea 0, 1, 2, ... k .

6.1 La aparición del pequeño mundo

Sus orígenes se encuentran en la sociología de las redes sociales; Milgram (1967) fue el primero en observar y enunciar el concepto. La expresión “pequeño mundo” fue utilizada para significar el hecho que dos personas cualesquiera, A y B, seleccionadas al azar en un medio social dado, están conectadas vía una cadena limitada de intermediarios conocidos. El estudio de las distancias en las redes sociales comienza en los años 1950 y 1960 más o menos. Pero la investigación específica del fenómeno de pequeño mundo no comienza que hacia fines de los años 1960 (Wasserman and Faust, 1999, p. 53-54; Degenne and Forsé, 2001, p. 20-23).

Desde el punto de vista que nos interesa aquí, Watt (1999) revisó su formulación (ver en particular el capítulo 2), y mostró que la fuente de las dificultades teóricas y empíricas en el corazón de los estudios pequeño mundo reposaba en la afirmación que las solas redes cuyas propiedades estadísticas son analíticamente tratables, son aquellas que son completamente ordenadas (*lattice graphs or d-lattice*) o completamente aleatorias (i.e. los grafos aleatorios). Estos dos extremos opuestos comparten la característica esencial que su estructura local refleja, exactamente o estadísticamente, su estructura global, y que por consecuencia el análisis basado en el conocimiento estrictamente local es suficiente para captar las estadísticas de la red en su totalidad. Después de una revisión crítica de los estudios pequeño mundo en el análisis de las redes sociales, Watts (1999) llega a la conclusión que el fenómeno del pequeño mundo había sido definido sin considerar de una manera precisa cuales son las características

específicas que una red debe poseer para exhibir dicho fenómeno. Su observación es que la propiedad pequeño mundo aparece en una zona intermedia entre por una parte las redes completamente ordenadas y por otra las redes completamente aleatorias.

Tanto el artículo de Watts & Strogatz (1998) como el libro de Watts (1999) muestran que los grafos pequeño mundo ocurren en una gran variedad de redes físicas, biológicas, tecnológicas y sociales.

6.2 Propiedades y definición

Los grafos pequeño mundo (*small world graphs*) son aquellos que son fuertemente en racimos como un grafo regular, pero sin embargo el largo de los caminos entre cualquier par de nodos es corto como en los grafos aleatorios. Por lo tanto, para definir los grafos pequeño mundo de una manera rigurosa, las estadísticas más importantes que debemos considerar son el “coeficiente de racimo” (*clustering coefficient*) y el “largo del camino promedio” (*characteristic path length*).

- El coeficiente de racimo $C(G)$ caracteriza la proporción de vecinos de un nodo que también son vecinos entre ellos, promediado por todos los nodos (Watt 1999, p. 32-33)
- El largo del camino promedio $L(G)$ es la distancia del camino mas corto entre dos nodos, promediado por todos los pares de nodos (Watts 1999 p. 29).

Ahora estamos en posición de poder definir la propiedad pequeño mundo de un grafo. Se llaman grafos pequeño mundo los grafos que exhiben el largo del camino promedio $L(G)$ equivalente a los grafos aleatorios, esto es, $L \approx L_{random}(n,k)$, pero un coeficiente de racimo $C(G)$ mucho más grande, $C \gg C_{random} \approx k/n$ (Watt, 1999, p. 29, 32-33, 58, 100, 114).

6.3 El pequeño mundo en la Web

Broder et al. (2000) proponen una representación de la Web en cuatro sectores mayores: un núcleo central o corazón, un sector más arriba y otro más abajo del mismo núcleo, y los llamados zarcillos (*tendrils*). El núcleo central contiene los sitios más prominentes. Los sitios “más arriba” pueden alcanzar el núcleo pero no pueden ser alcanzados desde el núcleo. Los sitios “más abajo” pueden ser alcanzados desde el núcleo pero en cambio no puede alcanzar a éste. Los “zarcillos” contienen los sitios que no pueden alcanzar ni son alcanzados desde el núcleo. Cada uno de estos sectores de la Web tiene aproximadamente el mismo tamaño. El núcleo es más compacto (densidad). Esta sería la estructura global de la Web, y la propiedad pequeño mundo encontrada por (Albert et al., 1999, Adamic, 1999, Broder et al., 2000, Kleinberg and Lawrence, 2001) se realizaría solamente en el núcleo. El largo del camino promedio L de la Web sería entre 16 y 19 o 20, y el coeficiente de racimo (o *clustering*) del orden de 0,11 a 0,20 en estos estudios.

De acuerdo con el estudio de Broder et al (2000), la estructura macroscópica de la Web es considerablemente más compleja, en el sentido que una gran cantidad de páginas no se encuentran conectadas, y además un número significativo no puede ser alcanzado que sólo mediante caminos que pasan por centenares de paginas intermediarias.

7. Centralidad, centralización y densidad

Aquí vamos a ver que las redes se caracterizan además por las propiedades estructurales de centralidad y centralización, y que estas nociones se declinan de acuerdo con los conceptos de grado, intermediación y proximidad. Además toda red posee una densidad. Estos son conceptos muy utilizados en el análisis de las redes sociales, y que podemos ampliar al análisis de redes en general. Por lo esencial, esta sección se basa en el artículo pionero de Freeman (1979) y en las presentaciones de Wasserman & Faust (1999, capítulo 5) y de Degenne & Forsé (2001, capítulo 6).

Notemos que los conceptos de centralidad y de centralización no han sido utilizados en los estudios de la Web a los cuales me he referido en las secciones anteriores, y que ahora proponemos para el análisis de las redes tecnológicas de la comunicación (Internet, Web) y las redes de conocimiento (ver sección 8)

Recordemos que el método de palabras asociadas (Courtial, 1990; Callon et al 1993), en el análisis de la información científica, desde un comienzo se ha servido de estas nociones de centralidad y densidad, aún cuando la definición no sea la misma, para caracterizar las clases o clusters, y así disponer los clusters sobre un espacio bidimensional cuyas coordenadas x e y son justamente la centralidad y la densidad; nada impide revisar por cierto el método de las palabras asociadas desde el punto de vista de la teoría de grafos, y por consecuencia de adoptar las definiciones de centralidad, centralización y densidad que son presentadas en esta sección.

7.1 Grado, intermediación, proximidad

Estos son conceptos que modelan una red y permiten así su análisis. Y los vamos a presentar rápidamente como introducción. En efecto, la centralidad y la centralización se declinan en el análisis de redes sociales en función de estos tres conceptos dando así origen a la definición y uso de tres clases de centralidad y de centralización.

El *grado* es el número de lazos de un actor en la red, la *intermediación* significa que un actor se encuentra entre otros dos actores en la red, la *proximidad* es la distancia entre un actor y resto de la red. Estas son tres propiedades estructurales características de los miembros de una red. La elección de un atributo estructural particular y su medida asociada depende de lo que busca analizar en la red: si se trata de evaluar la capacidad de comunicación de que un actor dispone en la red, la medida basada en el grado se impone; si nos interesamos por el control de la comunicación, la medida apropiada es la intermediación; si se trata de la independencia de un actor ella conduce a la elección de la medida basada en la proximidad.

7.2 Centralidad

El concepto de centralidad se refiere a la posición de los nodos en las redes, y la centralización al conjunto de la estructura de una red. Todas las medidas de centralidad y centralización asignan el valor más elevado a la estructura en estrella o rueda, y el valor mas bajo al grafo completo en donde todas las aristas posibles están presentes, puesto que todos los nodos en ese grafo son homogéneos en todos los aspectos.

La manera según la cual la centralidad sea definida depende de la medida como vemos en la tabla 1, y las conclusiones dependen de lo que se desea obtener en la aplicación empírica. Si

nos interesamos por un actor (representado por un nodo de la red) y los efectos de su posición estructural en ella, se hace necesario conocer su centralidad. Los individuos centrales ocupan una posición privilegiada en los intercambios, en particular por comparación a aquéllos que son rechazados a la periferia, ellos son los nodos más significativos de la red y es razonable pensar que esto se traduce en términos de poder.

Tabla 1: Centralidad. Las definiciones matemáticas y por consecuencia de cálculo tal como estas nociones son utilizadas en el análisis de redes sociales. Notación: C es la centralidad que es función de un nodo específico n_i , $i = 1, 2, \dots, N$, los índices G, I, P señalan el tipo de C , $C(n_i^*) = \max_i C(n_i)$, esto es, el máximo valor de un nodo i entre los N nodos de la red.

Grado	$C_G(n_i) = \sum_{j=1}^N g(n_i);$	$C_G^{Nor} = \frac{C_G(n_i)}{N-1}$
Intermediación	$C_I(n_i) = \frac{\sum_{j < k} N_{jk}(n_i)}{N_{jk}};$	$C_I^{Nor} = \frac{C_I(n_i)}{[(N-1)(N-2)/2]}$
Proximidad	$C_P(n_i) = \left[\sum_{j=1}^N d(n_i, n_j) \right]^{-1};$	$C_P^{Nor}(n_i) = \left[\frac{N-1}{\sum_{j=1}^N d(n_i, n_j)} \right]^{-1}$

7.2.1 Centralidad de grado

La centralidad de grado es la más simple y cerca de la intuición, ella privilegia el punto de vista local y mide la actividad o la capacidad de comunicación o intercambio de cada nodo dentro de la red, no tomando en cuenta su capacidad a controlar estas comunicaciones. El grado de un nodo es útil como índice de su potencial de comunicación.

7.2.2 Centralidad de intermediación

Podemos también pensar que un individuo (nodo o sitio de la Web) es del todo más central si depende menos de los otros en la comunicación con el conjunto de la red, en otras palabras que no necesita de relé para transmitir sus mensajes. Desde este punto de vista, un nodo es central en la medida que es independiente; independencia y control aparecen aquí como sinónimos. Para dar cuenta de este hecho dos medidas de centralidad son posible, la intermediación y la proximidad.

La intermediación es la frecuencia con la cual un nodo (o actor) se encuentra entre un par de otros nodos en el camino más corto o geodésica conectando a ellos. Un nodo cuyo lugar se encuentra en el camino de comunicación entre otros nodos exhibe un potencial de control de su comunicación. La idea que Freeman (1979) sugiere es que un individuo puede muy bien estar ligeramente conectado a los otros (es decir, centralidad de grado baja) y sin embargo ser un intermediario esencial en los intercambios. Como Degenne and Forsé (2001) observan, más este individuo (nodo) es útil o puede ser usado como intermediario por todos los otros miembros de la red, más está en posición de controlar la comunicación o ser independiente de los otros para comunicar. Un tal individuo puede más fácilmente influenciar el grupo filtrando o distorsionando la información en circulación, este individuo esta también en mejor posición para asegurar la coordinación de la red, de todos estos puntos de vista, ocupa una posición central. Para dar cuenta de esta posición la medida pertinente es la centralidad de intermediación.

7.2.3 Centralidad de proximidad

Una tercera manera de medir la centralidad de un individuo (nodo) consiste en medir su grado de proximidad con respecto a todos los otros individuos. Se trata de una medida más global utilizando no solamente las conexiones de un individuo a su vecindario (*neighbourhood*), sino que su proximidad con el conjunto de los miembros de la red.

Este punto de vista de la centralidad, que un nodo es próximo a todos los otros nodos del grafo, está relacionado también con el control de la comunicación, pero como Freeman (1979) subraya, de una manera algo diferente que la centralidad basada en la intermediación, aquí un nodo es visto como central en la medida que puede evitar el posible control de los otros. Una posición central es aquélla que no es dependiente de los otros como intermediarios del mensaje. Aquí las palabras centralidad e independencia pueden aparecer como equivalentes.

Una última observación, la centralidad en términos de grado, proximidad e intermediación tiene que ver con la definición de prominencia, y la identificación de los actores más importantes (o grupos de actores) de una red. En cambio, la densidad vista como medida de la cohesión conduce a la definición de cohesividad (*cohesiveness*), y la identificación de subgrupos cohesionados de actores dentro de una red.

NOTA: Aún queda por considerar: 1) la centralidad de información (ver Wasserman & Faust, 1999, p. 192-198), mientras que la intermediación se basa en el camino geodésico, la centralidad de información se centra en cómo la información puede seguir diversos caminos, ponderados por la fuerza del lazo y de la distancia; así como también 2) la reformulación de la centralidad de intermediación en un grafo ponderado (Freeman et al, 1991)

7.3 Centralización

Esta es la significación del término centralidad cuando es aplicado a toda la red: una red puede ser centralizada o descentralizada. Al igual que por la centralidad, existen tres medidas de la centralización y cada una corresponde a una de las propiedades utilizadas para definir la centralidad de los nodos de la red (grado, intermediación y proximidad). Se considera que una centralización de grado fuerte es una indicación de comunicación activa entre todos los miembros de la red, mientras que una centralización de fuerte proximidad o intermediación traduce el hecho que un número pequeño de actores controla esta comunicación.

Tabla 2: Centralización. Definiciones de las categorías de centralización de una red. La notación es la misma que la tabla 1, solo que C es aquí la centralización.

Grado	$C_G = \frac{\sum_{i=1}^N [C_G(n^*) - C_G(n_i)]}{\max \sum_{i=1}^N [C_G(n^*) - C_G(n_i)]} ; \quad C_G^{Nor} = \frac{\sum_{i=1}^N [C_G(n^*) - C_G(n_i)]}{[(N-1)(N-2)]}$
Intermediación	$C_I = \frac{2 \sum_{i=1}^N [C_I(n^*) - C_I(n_i)]}{[(N-1)^2(N-2)]}$
Proximidad	$C_P = \frac{\sum_{i=1}^N [C_P^{Nor}(n^*) - C_P^{Nor}(n_i)]}{[(N-2)(N-1)]/(2N-3)}$

7.4 Densidad

En su forma más general, la densidad de una red es la proporción de lazos existentes comparado a los lazos posibles. La definición intuitiva de densidad es entonces el número de relaciones efectivas R dividido por el número de relaciones posibles, y como se ha visto en la sección 5, el número de lazos posibles es igual a $N(N-1)/2$ cuando los lazos no son orientados, y si los lazos son orientados, $N(N-1)$. Existen tres medidas de la densidad de acuerdo con los tipos de grafos, tal como se encuentra en la tabla 3.

Tabla 3: Densidad. Medidas de acuerdo con los tipos de grafos

Grafo (G) y subgrafos	$D(G) = \frac{R}{N(N-1)/2} = \frac{2R}{N(N-1)}; D(subG) = \frac{2R}{N(N-1)}$
Grafo orientado (Go)	$D(Go) = \frac{R}{N(N-1)}$
Grafo ponderado (Gp)	$D(Gp) = \frac{\sum v_k}{N(N-1)}$

En las redes desigualmente conectadas, la densidad hace posible medir las áreas más o menos ligadas o enlazadas de la red, y detectar el vecindario de un nodo dado, como los grupos y “*cliques*”, los cuales son reconocidos por su cohesión de acuerdo con la proporción de relaciones entre los nodos. Una noción análoga es el *clustering* tal que Watts (1999) lo ha definido.

8. Modelos de análisis de redes de conocimiento

¿Cómo pasar del análisis de la red tecnológica es decir la Web al análisis de las redes de conocimientos que se encuentran presentes en ella? Como hemos dicho, las redes de conocimiento (R_C) que nos interesan son aquellas que se encuentran al interior de la estructura de la Web (R_{TC}). Y este conocimiento lo llamaremos entonces e-conocimiento, dado que se encuentra en los documentos electrónicos, disponibles en la Web o accesibles por medio de esta red tecnológica. Además, cuando hablo de conocimiento no me estoy refiriendo al conocimiento como fenómeno psicológico o mental, hablo del conocimiento como entidad objetiva existiendo en los documentos (papel o electrónico, digital), es decir, que se encuentra en el lenguaje escrito de las publicaciones. En otras palabras, hablo de lo que Popper (1979) llama “conocimiento objetivo” desde el punto de vista de una “epistemología sin sujeto” (“epistemology without knowing subject”).

En esta sección me referiré a los modelos de análisis que han sido algorítmicamente implementados e incorporados en la R_{TC} , y que provienen de la bibliometría o cienciometría y cuyo objetivo es el de identificar redes de conocimiento y competencias: el modelo de las citas y el modelo de la co-ocurrencia (co-citas y palabras asociadas), y además el modelo de las redes neuronales de auto organización.

8.1 Redes de conocimientos

Basándonos en el concepto general que toda red es a la vez una estructura, un contenido y un uso, hemos distinguido (en la sección 2) tres tipos de análisis de la red tecnológica de comunicación a través la cual contenidos de información y de conocimiento se difunden y

están disponibles. El análisis de contenido se refiere aquí a la red de conocimientos existentes en la red tecnológica de comunicación (Internet, Web), y que podemos expresar mediante la fórmula (R_{TC} (R_C)). La cuestión es entonces de saber cómo analizar las redes de conocimientos R_C en la red tecnológica R_{TC} , es decir, Internet y la Web.

Observemos que la R_{TC} en donde se encuentra la R_C , presenta las mismas propiedades estadísticas que la bibliometría y la cienciometría han descubierto en la literatura científica, desde ya hace bastante tiempo, y enunciado como las leyes y distribuciones hiperbólicas de Lotka, Bradford y Zipf (a propósito respectivamente de los autores, de las revistas y de las palabras al interior de un texto). Es así que siguiendo el modelo de la citación para analizar la Web, Rousseau (1997) encuentra que la distribución de los “in-links” (de 343 sitios) es una distribución de Lotka, y Cui (1999) aplica la distribución de Bradford para determinar cuáles son los sitios Web mas centrales en medicina. Al mismo tiempo, y entre otros analistas que no vienen del campo de la ciencia social, Lada Adamic estima que la distribución de visitas a los sitios de la Web presenta la forma de la ley Zipf-Pareto⁷. La conclusión sería entonces que las redes de conocimientos presentarían igualmente una estructura de ley de potencia (Bradford, Lotka, Zipf, Pareto). Sin embargo, debemos matizar esta afirmación, y tener en cuenta la observación de Pennock et al (2002) que “entre las páginas Web del mismo tipo, el conjunto de la distribución de enlaces entrantes se desvía fuertemente de una ley de potencia, exhibiendo aproximadamente una forma log-normal”. Pero lo que me parece aquí más importante es que podemos representar y analizar las redes de e-conocimiento sirviéndonos de la teoría de grafos: las páginas Web (en donde se encuentran los conocimientos en su forma de lenguaje escrito) son los nodos y los hiper-enlaces son las relaciones.

8.2 Modelo de la citación

La red que constituyen los artículos científicos a través las citaciones fue subrayada en los años 1960 primero por Garfield (1963) y enseguida por Price (1965), es decir, cada publicación representa un nodo y las citaciones son las relaciones entre estos nodos; al interior de este tipo de red (o grafo) es que el conocimiento científico se desarrolla y se puede medir estadísticamente (Price, 1986), o estudiar con ayuda de la teoría de grafos siguiendo el ejemplo del análisis de redes sociales. Lo notable es que los analistas de la Web ajenos a esta tradición del análisis de citaciones lo han redescubierto, estimándolo como un modelo que puede aplicarse igualmente al análisis de la Web, considerando entonces los documentos de la Web y sus enlaces hipertexto (Kleinberg, 1998, 1999, Kleinberg & Lawrence 2001, Chakrabarti 2003). Digamos que la metáfora de los enlaces como citaciones puede ser válida, como también la analogía, pero no la asimilación de los enlaces a las citaciones. La observación de uno de los primeros a referirse al análisis de citaciones, Kleinberg (1998), va en este sentido: el análisis de los enlaces tiene al mismo tiempo relación y contrasta con el análisis de citaciones (véase, además, Thelwall 2003, quien insiste sobre las diferencias de motivación y naturaleza).

8.2.1 Centros de actividad (*hubs*) y autoridades

En lo que se refiere al acceso a los conocimientos contenidos en la Web a través del análisis de los enlaces de la Web, se considera que este análisis de la estructura de los enlaces permite definir tópicos o temas de conocimiento. En el sentido que una densidad elevada de enlaces entre un conjunto delimitado de páginas (o sitios) estaría indicado que ellos pueden estar

⁷ <http://www.hpl.hp.com/research/idl/papers/ranking/ranking.html>

temáticamente relacionados. Es el caso de una colección de las páginas llamadas “hubs”, y significadas por la cantidad de enlaces salientes, ligadas así a una colección de páginas que aparecen como “autoridades” acerca de un tópico común (Kleinberg 1998, 1999, 2001). Desde el punto de vista de la documentación, Ingwerse & Björneborn (2004) observan que subyacente a la distinción propuesta por Kleinberg (1999) entre “hubs” y “authorities”, estaría la idea de documentos altamente citados (*authorities*) y de artículos de síntesis o *reviews* (*hubs*).

En una cierta medida, toda página de la Web es a la vez “hub” y “authority” según grados diferentes. De manera que cada página p tiene dos medidas diferentes, un “hub score” $h(p)$ y un “authority score” $a(p)$. Las definiciones cuantitativas de “hub” y “authority” son recursivas, es decir, el “authority score” de una página Web es proporcional a la suma de “hub scores” de las páginas enlazándose a ella, e inversamente su “hub score” es proporcional a los “authority score” de las páginas con las cuales ella se enlaza.

El análisis de la estructura de los enlaces de la Web puede también ayudar a definir grupos sociales o comunidades de interés para los usuarios de la Web (Kleinberg and Lawrence, 2001). Por ejemplo, Kumar et al. (1999, 2000) consideran que una comunidad de conocimiento puede ser identificada mediante la detección de un conjunto entrelazado de “hubs” (>> enlaces emitidos) y “authorities” (>> enlaces recibidos).

8.2.2 Web comunidades

Se ha llamado “comunidad Web” (*Web community*) una colección de páginas Web en la cual cada página miembro tiene más hiper-lazos al interior de dicha colección que con el exterior. Flake et al (2002) generalizan esta definición “a la identificación de comunidades con tamaños y conectividad variables” y para ello proponen un enfoque algorítmico fundado en la teoría de grafos. Esta es otra manera de identificar redes de e-conocimiento.

8.2.3 PageRank

El índice de popularidad o prestigio de las páginas Web de Google (Page & Brin 1998), puede ser adaptado para el análisis de las redes de e-conocimiento, se trata del célebre PageRank, o $PR(A)$. (Para una comprensión técnica, véase la sección 7.2 del manual de Chakrabarti (2003); y para una presentación pedagógica, ver “A Survey of Google’s PageRank” en eFactory⁸). En el marco del análisis de citaciones, los artículos más populares o prestigiosos son los más citados. Considerando los enlaces entre páginas como citaciones, el PageRank de una página A se define como $PR(A) = (1-d) + d (PR(T1) / C(T1) + \dots + PR(Tn) / C(Tn))$, es decir, la página A tiene $T1 \dots Tn$ páginas cuyos enlaces apuntan a ella, y que consideran como citaciones de la página A por las páginas $T1 \dots Tn$; y se utiliza d que es un factor de amortiguamiento (*damping*), es decir, la probabilidad que el navegador pase aleatoriamente de una página a otra. Este índice es entre 0 y 1, usualmente d es fijado en 0,85. Además puede decirse que $C(A)$ es el número de enlaces saliendo de la página A . Como vemos, el $PR(A)$ es una medida de cuanto es citada una página y que se estima como un indicador de importancia. En el fondo, el RankPage es una adaptación astuciosa del análisis de citaciones al gigantesco hipertexto de la Web, y que ahora puede igualmente utilizarse para cartografiar su estructura intelectual, para analizar la red de e-conocimientos contenida en la estructura de los enlaces hipertexto de la Web.

⁸ Disponible (1/03/2006): <http://pr.efactory.de/>

8.3 Modelo de la co-ocurrencia o asociación

Pasemos a considerar el modelo de la co-ocurrencia y que proviene también del campo de la cienciometría. Este modelo es la base del análisis de co-citaciones como también y por sobre todo del análisis de las palabras asociadas. Callon et al (1993) insisten que mientras el método de co-citaciones realiza una clasificación temática indirecta, el método de las palabras asociadas (*coword analysis*) aparece como mucho más apto para captar directamente el contenido de los documentos, es decir, más adaptado para trabajar el lenguaje escrito de los documentos científicos (como también otros). En la lingüística encontramos el triángulo “palabra, concepto y mundo”, es decir, la palabra es un signo (fonético o gráfico) cuya significación es el concepto que envía al referente, cosa u objeto del mundo (es decir, la realidad extra lingüística percibida). Este triángulo opera en el lenguaje científico. Y lo que aquí llamamos conocimiento se encuentra en el concepto al cual hace referencia la palabra.

En realidad, el análisis de la Web puede basarse en los enlaces (*link-based*) o en los documentos y paginas (*text-based*). El método de las palabras asociadas sería mucho más adaptado para desarrollar un “text-based” análisis de la Web, puede aplicarse directamente al análisis de documentos electrónicos, lo que cambia no es el método en si mismo sino la naturaleza del objeto de análisis (en lugar de los tradicionales datos bibliográficos, se trata ahora de paginas Web, de textos electrónicos).

8.3.1 Método general

El método general y que podemos llamar “co-(x) análisis”, siendo x cualquier ítem (citaciones, palabras claves, sitios y enlaces, o bien usuarios), se compone de cuatro etapas: 1) la transformación de los datos en una matriz de asociación; 2) la utilización de un índice o coeficiente de asociación; 3) la aplicación de un método de clasificación automática (la partición de la red de asociaciones en clases o clusters); 4) la representación gráfica de las clases en un mapa (espacio de visualización y de análisis).

Apoyándonos en nuestra propia experiencia, el modelo de la co-ocurrencia puede aplicarse, como vamos a mostrar, de acuerdo con un proceso en cascada: 1) a la estructura de los enlaces, para identificar los sitios claves en un sector de la Web, como co-sitio análisis; 2) al contenido de las páginas de los sitios claves así identificados, como método de las palabras asociadas; 3) al uso de esos documentos, usuarios y objetos utilizados, como análisis de co-uso. El método es el mismo lo que cambia es el objeto al cual se aplica. No son más que modalidades de aplicación del mismo enfoque fundado en el modelo de la co-ocurrencia.

Dos atributos x_i y x_j están relacionados formando una pareja x_{ij} , cuando ellos se refieren al mismo dato d , de la misma manera dos datos, d_i y d_j , están relacionados formando una pareja d_{ij} , cuando un mismo atributo x se refiere a esos dos datos. Así, cada atributo será descrito por el conjunto de co-datos d_{ij} :

- $x_i = [d_{ij}]_{i=1,m; j=1,n}$; siendo $d_{ij} \in [0,1]$; y m = el numero de atributos, n = el numero de datos.

Y a su vez cada dato por un conjunto de co-atributos x_{ij} :

- $d_i = [x_{ij}]_{i=1,n; j=1,m}$, siendo $x_{ij} \in [0,1]$; y m = el numero de atributos, n = el numero de datos.

Enseguida se utiliza un coeficiente de asociación, para normalizar las co-ocurrencias. Y luego se procede a la clasificación automática pudiéndose así generar clusters de atributos y/o clusters de datos. En las aplicaciones a las que aquí me refiero como ejemplos de análisis de la Web, el coeficiente de asociación que se utiliza es el coeficiente de equivalencia:

- $E_{ij} = C_{ij}^2 / F_i \times F_j$, es decir, la co-ocurrencia C de i y j al cuadrado dividida por la frecuencia F de i y j .

En cuanto al método de clasificación automática, se aplica un algoritmo fundado en el método de clasificación jerárquica ascendente, de acuerdo con la regla del simple enlace (“single linkage”).

8.3.2 Análisis de co-sitios

Aquí se trata de clasificar en clases o clusters los sitios en función de sus enlaces. Dos sitios, i y j , están asociados no por sus enlaces directos sino que por los enlaces enviados desde otros sitios, x_1, x_2, \dots, x_n . Como lo muestra la figura 2, hay dos opciones posibles, aquí se aplica la primera y por eso hablamos de co-sitio análisis y no de co-enlace análisis. El valor numérico de la asociación $a(i,j)$ es igual al número de sitios al origen de los enlaces. Este valor absoluto es normalizado mediante el uso del coeficiente de equivalencia E_{ij} . Enseguida se procede a la clasificación de los sitios así asociados, lo que da como resultado un cierto número de clases o clusters.

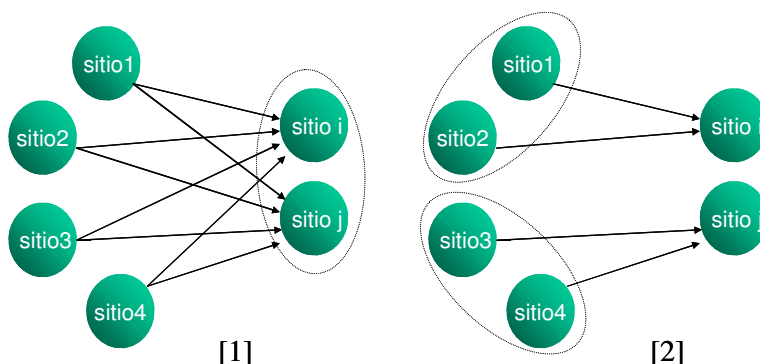


Figura 2: Modalidades de calcular las asociaciones. [1] el enfoque co-sitios, $C(i,j)$, si los sitios i y j ambos son enlazados por 1, 2, ..., n otros sitios se dice que están asociados entre ellos y [2] el enfoque co-enlaces, $C(1,2)$ y $C(3,4)$, dos sitios se consideran asociados si los ambos tienen enlaces con un mismo sitio (sea i o j). [1] Corresponde en bibliometría al paradigma de co-citaciones, si ambos artículos A y B son citados por el artículo C, pueden considerarse asociados, aunque no se citen directamente, y también de las palabras asociadas, si las palabras claves A y B indexan el artículo C pueden considerarse asociadas; [2] corresponde al paradigma que se conoce como “bibliographic coupling” que asocia dos artículos A y B que citen a un mismo artículo C.

Esta manera de proceder fue aplicada por la primera vez en Polanco et al (2001), sobre una muestra de 791 sitios de universidades perteneciendo a los 15 países miembros de la Unión europea (UE) a esa fecha, entrelazados por 5.308.204 enlaces, y representando 12.595.809 páginas. Estos 791 sitios fueron agrupados en 37 clusters.

Más tarde estos 37 clusters fueron revistos desde el punto de la teoría de grafos (Polanco 2002), lo cual permite de representar la red de clusters (subredes) por un grafo orientado y ponderado en lugar del mapa bidimensional, y de analizarlos de acuerdo a los tipos de relaciones y los tipos de nodos, como se ha expuesto mas arriba en las secciones 5.4 y 5.5.

Las figuras 3 y 4 representan dos espacios de representación completamente diferentes. Siendo que la relación de co-ocurrencia no es una relación orientada, y sobre la cual esta fundada la relación entre clusters, el hecho de venir orientadas (en la figura 4) es una consecuencia del algoritmo de clasificación aplicado por el programa SDOC que utilizamos.



Figura 3: Mapas del análisis de co-sitios y de palabras asociadas (co-words). A izquierda, el mapa de los 791 sitios académicos agrupados en 37 clases o clusters basado en la co-ocurrencia de sitios (ver figura 2 [1]). A derecha, el mapa de las 1.353 páginas del sitio del Laboratorio de Computación de la Universidad de Cambridge, agrupadas en 54 clases o clusters basado en la co-ocurrencia de palabras claves. En los dos casos, cada clase o cluster es un grafo cuyos vértices o nodos son los sitios o las palabras claves y cuyas aristas son las relaciones de co-ocurrencia de estos ítems. Además, por efecto del algoritmo de clasificación, dos vértices asociados por una relación de co-ocurrencia pueden encontrarse en dos clases diferentes, lo cual da origen a relaciones entre las clases. Este conjunto de clases y de relaciones inter-clases constituye una red que puede ser representada a su vez por un grafo como lo muestra la figura 4.

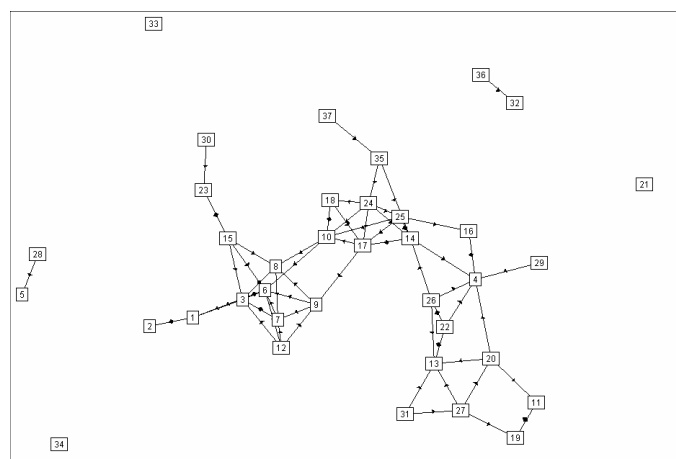


Figura 4: El grafo de la red de clases o clusters. Aquí la representación no es bidimensional sino que ella muestra la red que forman los 37 clusters de acuerdo con sus interrelaciones (inter-clases). Cada clase es un nodo enumerado y la orientación de las relaciones están señaladas por puntas de flechas en medio del arco. El valor de cada arco está dado por el índice E_{ij} . La lectura de este mapa debe hacerse de acuerdo con los principios expuestos en la sección 4 y 5. La centralidad y la densidad se definen y pueden calcularse de acuerdo con lo dicho en la sección 6.

Las redes de co-sitios pueden ser generadas para sectores específicos de la Web, y sometidas a una clasificación automática (*clustering*). Los clusters de co-sitios proporcionan un nuevo modo de estudiar la estructura de la Web y su desarrollo. En general, un método de clasificación trata de organizar (agrupar) los objetos dentro de grupos relativamente homogéneos. Los clusters representan grupos (o racimos) de sitios de la Web altamente similares. La similitud está basada en este caso en el grado de asociación (co-ocurrencia) que existe entre dichas entidades, y el cual está dado por su valor de E_{ij} (en una escala entre 0 y 1).

8.3.3 Sobre páginas Web

El mismo principio de $co(x)$ análisis, bajo su forma tradicional de análisis de las palabras asociadas (*co-word análisis*), lo hemos aplicado tal cual sobre paginas de la Web, en junio del 2003 (ver sección 8.4.3). En lugar de documentos tradicionales o datos bibliográficos, se trata ahora de páginas de la Web y extraídas de un sitio, el Computer Laboratory de la Universidad de Cambridge⁹. En total, 1.353 páginas indexadas por 1.230 palabras claves después de una validación humana y habiendo sido extraídas directamente de estas mismas páginas por un analizador lexicográfico. A partir de la matriz de datos $D(N,P) = D(1.353 \times 1.230)$ se aplicó el procedimiento estándar del método de las palabras asociadas. Resultando 54 clases o clusters que se consideran como indicadores temáticos. Las 54 clases están representadas en un mapa $M(C,D) = M(\text{centralidad} \times \text{densidad})$ como muestra la figura 3, al igual que las 37 clases de sitios académicos. En presentado estos dos mapas uno al lado del otro, en la figura 3, estamos sugiriendo el encadenamiento del proceso de análisis de la estructura de un campo de la Web y enseguida del contenido de un sitio en particular, y que puede extenderse a todos los sitios de una misma disciplina científica y presentes en las clases producidas por el análisis de co-sitios.

8. 3.2 Utilización del conocimiento

El propósito es aquí de analizar la utilización de los conocimientos disponibles como documentos accesibles vía la Web; se trata del análisis de la utilización y de la co-utilización y que hemos expuesto en detalle en nuestro artículo de *Scientometrics* (Polanco et al 2006). Este análisis explota la información contenida en los “log-files”. Y comprende dos tipos de enfoque. El primero consiste en la producción de una información estadística del uso, esto es, por una parte de los documentos utilizados (es decir, consultados), y por otra del comportamiento de los usuarios, sea en la búsqueda de información, sea en la realización de transacciones comerciales (como por ejemplo pasar una orden de pedido a propósito de un documento de interés). Un sistema que hemos llamado Miri@d realiza este análisis estadístico y además permite la realización de un análisis en línea.

Además, dos factores de impacto de uso son utilizados. Un *factor de utilización*, $\Sigma dc / R$, que es la proporción de los artículos de una revista consultados por usuarios de la Web en un periodo de tiempo determinado (dc), y el total de artículos publicados en esta revista durante el mismo periodo R . Y un *factor comercial*, $\Sigma dp / R$, que es la proporción de artículos de una revista pedidos por un cliente durante un período de tiempo determinado (dp), y el número total de artículos publicados en esta revista durante el mismo período. A estos dos factores pueden aún agregarse otros dos *factores de obsolescencia* de la información utilizada, o siendo el objeto de una transacción comercial (de compra y venta).

⁹ <http://www.cl.cam.ac.uk/>

El análisis de co-uso está basado en la medida de la co-ocurrencia. El objetivo de este análisis es identificar y visualizar los centros de interés en el uso de ciertos conocimientos o áreas problemáticas. En la noción de uso encontramos por una parte el usuario y por otra el objeto utilizado. Por lo tanto el análisis de co-uso comprende estas dos dimensiones. En este caso se trata de un ejemplo real de 410 usuarios habiendo solicitado 3.914 documentos (artículos científicos)

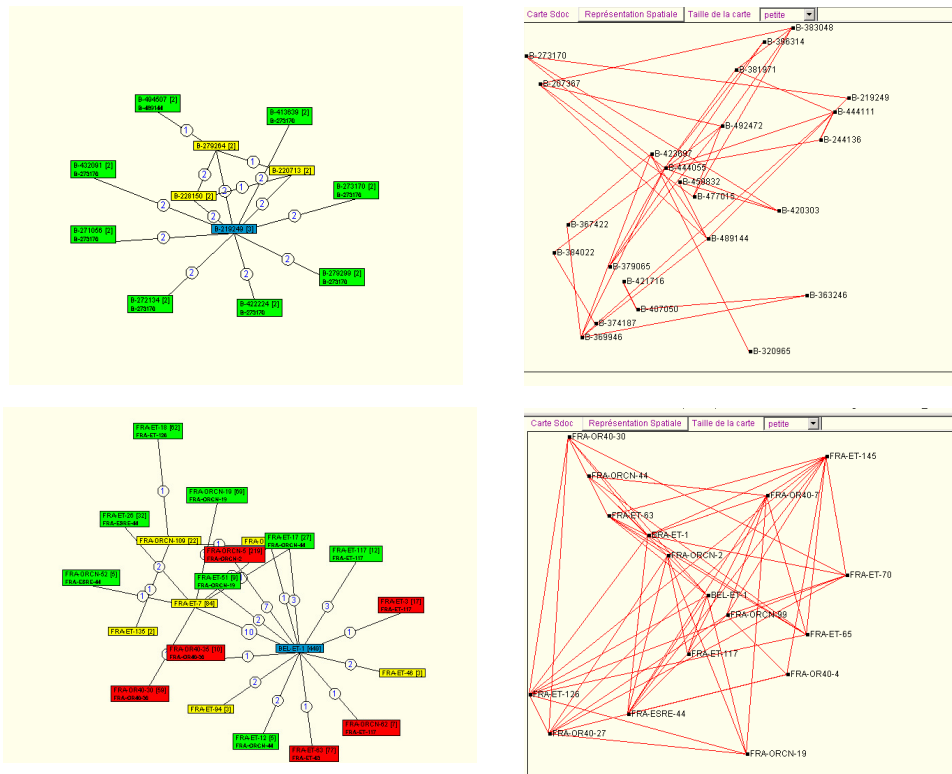


Figura 5: Grafos de clases o clusters y mapas. El grafo superior es una clase de documentos solicitados, el inferior una clase de usuarios. Sobre las asociaciones, en un círculo, se indica el número de usuarios al origen de las asociaciones, en el grafo superior, y en el inferior el número de documentos asociando dos usuarios. Por razones de confidencialidad usuarios y documentos están cubiertos por un número de código. Los mapas son del tipo $M(C,D)$, es decir, que las coordenadas X e Y corresponden a la centralidad y densidad. En el mapa superior se ve la posición de las 24 clases de documentos solicitados, y en inferior de las 17 clases de usuarios. Además, en ambos mapas se muestran todas las relaciones entre las clases, lo que hace visible las estructuras de las redes de clases.

Como hemos explicado (Polanco et al 2006), dos usuarios u_i y u_j están relacionados formando una pareja u_{ij} , cuando ellos se refieren al mismo documento d_i , de la misma manera dos documentos, d_i y d_j , están relacionados formando una pareja d_{ij} , cuando un usuario u_i se refiere en su interrogación a esos dos documentos. Así, cada usuario será descrito por un vector de co-documentos d_{ij} , y a su vez cada documento utilizado por un vector de co-usuarios u_{ij} . Estas asociaciones son ponderadas por el coeficiente E_{ij} . Y luego la clasificación automática genera las clases de usuarios y las clases de documentos utilizados. En nuestro ejemplo, 17 clases de usuarios y 24 clases de documentos solicitados. La figura 5 presenta dos ejemplos de grafos representando una clase de documentos y una clase de usuarios, y muestra

al mismo tiempo los mapas en donde se observan todas las relaciones entre clases formando la red en la cual cada clase es un nodo.

En conclusión, esta sección 8.3 viene de mostrar como un método general fundado en la co-ocurrencia puede aplicarse a la *estructura*, al *contenido* y al *uso* de la Web con el objetivo de analizar las redes de conocimientos. En la sección siguiente será cuestión de un modelo de red neuronal artificial aplicada al análisis de la Web.

8.4 Modelo de los mapas de auto organización

Como hemos visto (en las secciones 3 y 4), la Web es una red que se auto organiza sin necesidad de una autoridad central. El modelo de análisis que vamos a considerar aquí obedece a este principio de auto organización. En efecto, a partir del ejemplo de la auto organización en la red neuronal del cerebro humano, Kohonen formaliza entre 1981 y 1982 un algoritmo conocido como “*self-organizing map*” (SOM).

Quien dice mapa está significando una cierta estructura que hace posible la visualización incluso de fenómenos como la información, el conocimiento. En el campo de la extracción de conocimientos (“*knowledge discovery*”), se ha observado que para ciertas tareas de análisis, la visualización es el solo instrumento que es necesario para resolver un problema o confirmar una hipótesis, aún cuando las estructuras visuales no se piensen como un tipo de análisis (Brachman & Anand 1996). Los mapas pueden ser considerados como “*visualization-based analysis tools*”, es decir, no solo son medios de visualización sino que también constituyen herramientas de análisis. Recordemos aquí la definición de visualización propuesta por Card et al (1999): el empleo de representaciones visuales, basadas en programas informáticos interactivos, de datos abstractos, es decir sin dimensión física, con el objetivo de amplificar la cognición humana.

8.4.1 SOM

SOM (Kohonen, 1997) es un método de clasificación automática no supervisada (*clustering*) que clasifica directamente sobre un mapa auto-organizativo, y cuyo método de aprendizaje es competitivo. Notemos que el comportamiento auto-organizativo está estrechamente ligado con el método de aprendizaje competitivo. En el caso de SOM, el mapa es una grilla regular de neuronas o nodos, cada uno representado una clase o cluster. Lo que se llama en informática metafóricamente “neurona” o para ser más exactos “neurona artificial”, es en realidad un autómata simple operando al interior de una red.

El algoritmo de los mapas auto-organizados de Kohonen es bastante similar al método de *k*-medias (*k-means*), esto es, agregación alrededor de centros móviles, iniciación aleatoria, afectaciones a los centros de distancias mínimas, obtención de mínimos locales, pero la diferencia es que el algoritmo SOM conduce a una representación plana (a un orden espacial). El principio de base es la consideración que nuestra propia organización del conocimiento se crea durante el aprendizaje mediante mecanismos que promueven la auto-organización en un orden espacial (Kohonen 1997 p. 79-83)

El algoritmo consiste básicamente en: 1) la selección del nodo ganador de acuerdo con un aprendizaje competitivo, 2) la actualización de los valores del nodo ganador y de sus nodos vecinos. Lo mismo dicho de una manera más técnica en la tabla 4.

La aplicación de SOM a la Web es conocida como WEBSOM¹⁰ (Kohonen et al 1999). En este contexto, SOM es utilizado para organizar automáticamente grandes colecciones de documentos de la Web en mapas. El mapa constituye un paisaje de documentos en donde los documentos similares aparecen próximos como puntos de una grilla regular. Este paisaje puede ser etiquetado con palabras claves (o descriptores) automáticamente establecidas. Mediante un mecanismo interactivo HTML el mapa sirve para navegar en la colección de documentos y realizar búsquedas (ver la tesis de Arista Lagus, 2000)¹¹.

Tabla 4: Algoritmo SOM. Las operaciones principales de este algoritmo de aprendizaje competitivo (según Kohonen, 1997, p. 86-87).

-
- Sea $x(t) = \{x_1(t), x_2(t), \dots, x_N(t)\}$ el vector de entrada seleccionado al tiempo t , y $W_k(t) = \{W_{k1}(t), W_{k2}(t), \dots, W_{kN}(t)\}$ los pesos o valores del nodo k al tiempo t . La más pequeña de las distancias euclidianas define el nodo ganador k : $\|x(t) - W_k(t)\| = \min \|x(t) - W_k(t)\|$.
 - Después que el nodo ganador k es así seleccionado, los pesos de k y de los nodos vecinos (por ejemplo, dentro de un cuadrado o de un círculo alrededor del nodo ganador) son ajustados es decir sus pesos son modificados de acuerdo con la relación: $W_{ki}(t+1) = W_{ki}(t) + \alpha(t) \times h(t) \times [X_i(t) - W_{ki}(t)]$ para $1 \leq i \leq N$, siendo $\alpha(t)$ un índice de aprendizaje ($0 \leq \alpha(t) \leq 1$) que disminuye en el tiempo y converge a 0, y $h(t)$ la función de vecindad
-

Tabla 5: Generalización e Intercomunicación en Multi-SOM.

Generalización:

- $M \rightarrow M+1$
- $G(M+1)n = \frac{1}{4} \sum N(M)n$

Sea un mapa M cuyas dimensiones son $n \times m$ ($n, m \geq 2$), la generalización produce un nuevo mapa $M+1$, cuyas dimensiones son $(n-1) \times (m-1)$, gracias a la fórmula, en donde $N(M)n$ es el vecindario en el mapa M de donde proviene el nodo n del nuevo mapa $M+1$. Los datos originales son reafectados a los nuevos clusters del mapa $M+1$.

Intercomunicación:

- $A(S)i \rightarrow A(T)j$
- $A(T)j = f[g(A(S)i)]$

La activación de un nodo j del mapa T (*target*) resulta de la activación del nodo i del mapa de origen (*source*), de acuerdo con la fórmula, cuyos parámetros son las funciones f (probabilista o posibilista) y g (que es una “*bias function*”).

8.4.2 Multi-SOM

Multi-SOM es una extensión multi-mapas de SOM, de ahí su nombre, y fue desarrollado en el marco de un programa europeo, el programa EICSTES¹². Originalmente, la idea fue propuesta por Lamirel (1995). Multi-SOM permite, y esa es la innovación, de trabajar con varios mapas conectados entre ellos y cada uno representando un punto de vista sobre el objeto de estudio. Es decir, un mismo objeto puede ser analizado desde diferentes puntos de vista.

Un mecanismo de comunicación entre los mapas (o puntos de vista) permite de conectar y de observar las relaciones entre los diferentes mapas. Además cada mapa puede ser generalizado interactivamente, el sistema agrega una manera coherente las clases en clases más generales. En suma, Multi-SOM presenta tres instrumentos principales de análisis: el punto de vista, la generalización y la intercomunicación de los mapas. Así como también ciertas operaciones

¹⁰ Ver <http://websom.hut.fi/websom/>

¹¹ Disponible (5/03/2006): <http://www.cis.hut.fi/krista/thesis/>

¹² <http://www.eicstes.org/>

sobre las clases o nodos del mapa como poder las etiquetar, y ver la influencia de una clase (nodo) sobre las otras clases en el mapa, es decir su “zona de influencia”.

El *punto de vista* se define empíricamente por un subconjunto de palabras claves seleccionadas del conjunto total indexando una colección de documentos (que se supone representativos de un dominio científico determinado). La *generalización* es un proceso análogo a la inducción, esto es, a partir de las clases originales del mapa se calculan clases cada vez más genéricas, dando origen a niveles de abstracción. La comunicación entre mapas o *intercomunicación* es un mecanismo que activa relaciones entre los mapas, y permite de saber como una información particular en un mapa se refleja en los otros. Este mecanismo correspondería a una clasificación jerárquica basada en la noción de vecindario (*neighbourhood*).

8.4.3 Aplicaciones

Multi-SOM utiliza la versión estándar SOM_PACK¹³. Una interfase grafica en JAVA permite al usuario de consultar y trabajar sobre las aplicaciones¹⁴. Las aplicaciones sobre los datos de la Web tratan de captar y analizar la estructura académica de la Web en los países de la Unión europea, luego estudiar un campo del conocimiento científico, y en fin los proyectos presentes en las páginas del sitio de un laboratorio de investigación.

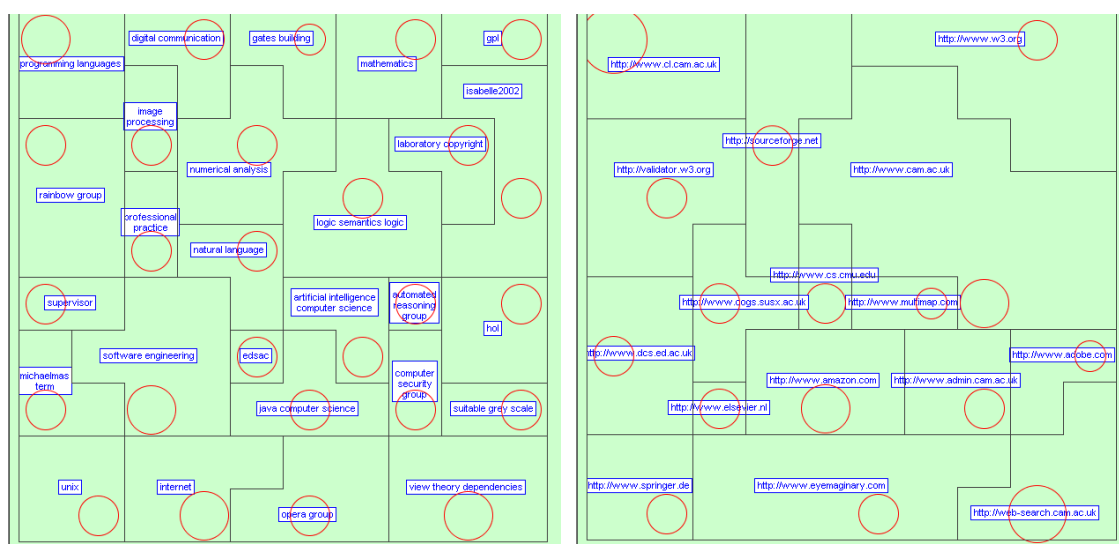


Figura 6: Mapas expresando puntos de vista. A izquierda el punto de vista de los proyectos de investigación (contenido) y a derecha el punto de vista de los sitios a los cuales se enlazan estos proyectos (estructura). Estos dos mapas están interconectados lo que permite pasar interactivamente de uno al otro.

En junio del 2003, Multi-SOM fue aplicado a una muestra de 1.736 sitios (URL), de un total de 427 universidades de la Unión europea, indexados por la categoría “ciencia y tecnología de la computación” del código UNESCO, de acuerdo con cuatro puntos de vista es decir mapas: 1) la distribución de los sitios por disciplina (de acuerdo con el código UNESCO), 2) la distribución de los sitios por país, 3) la red de enlaces entre los sitios, 4) la red de co-enlaces. Este estudio macroscópico fue seguido por un estudio microscópico sobre las páginas de un solo laboratorio de investigación al cual me he referido (en la sección 8.3.3), es decir, del Computer Laboratory of Cambridge University. Las 1.353 páginas HTML de este sitio fueron

¹³ Disponible libremente en www.cis.hut.fi solo para fines de investigación.

¹⁴ Las aplicaciones de Multi-SOM son disponibles (1/03/2006) en <http://eicstes.inist.fr/visualisation.html>

clasificadas y cartografiadas desde dos puntos de vista: 1) los proyectos o temas de investigación y 2) los sitios a los cuales estos temas envían enlaces. Como ya hemos dicho, las páginas fueron indexadas automáticamente por 1.230 palabras claves. Los enlaces saliendo del sitio fueron estandarizados conservando solo la raíz de la URL de los sitios enlazado¹⁵, un total de 1.912 sitios. La figura 6 expone los dos mapas reflejando los dos puntos de vista: los proyectos de investigación y los enlaces de estos mismos proyectos a otros sitios de la Web (y que podemos considerar como sus vecindarios).

9. Conclusión

En este ensayo he tentado de proporcionar un esbozo general acerca del análisis de redes. El problema que nos planteamos desde el comienzo es de saber *cómo* podemos conocer o analizar las redes (sociales, tecnológicas de la comunicación y de conocimiento), y *qué* es lo que hay que conocer y analizar en las redes. Para esta tarea, todos los conceptos propuestos se definen en términos matemáticos permitiendo realizar cálculos y medidas, y además pudiendo traducirse en algoritmos ejecutables por los computadores. Se ha visto que lo que hay que conocer y analizar en las redes son las estructuras y la dinámica (evolución, desarrollo, crecimiento), los contenidos y usos. Acerca de cómo podemos analizarlos, se ha insistido sobre el empleo de la teoría de grafos en lo que se refiere a la estructura; de la ley de potencia libre de escala en el análisis de las redes en tanto que sistemas complejos no lineales; y de la clasificación automática no supervisada (*clustering*) combinada con la teoría de grafos para el análisis de los contenidos y usos de la Web que, por cierto, es el ejemplo más evidente de una red tecnológica de la comunicación en la cual habitan las redes de conocimientos. En la última sección fueron abordados los modelos viniendo de la cienciometría y utilizados para el análisis de la estructura, del contenido y del uso de la Web. Se insistió particularmente sobre la base de nuestra propia experiencia en la aplicación del modelo de la co-ocurrencia, y concluimos con la aplicación reflexiva del principio de la red al estudio de la red vía los mapas de Kohonen en una versión multi-mapas.

Permítaseme, por último, las siguientes observaciones generales:

- El llamado paradigma de la información se desarrolla en el marco de una economía que Castells (1996) llama “capitalismo informacional”, en el sentido que sin las nuevas tecnologías de la información el capitalismo global no sería lo que es actualmente. Es decir, que la sociedad de la información o la sociedad del conocimiento no significa que hemos cambiado de sistema económico (o modo de producción como decía Marx). En una nota al fin de la sección 4 he insistido que las redes aparecen como sistemas competitivos sin ofrecer otra alternativa que “winner-take-all”.
- Con las nuevas tecnologías de la información y de la comunicación, habría aparecido la llamada “nueva economía” (Artus 2001), asimismo que una “economía del conocimiento” (Foray 2000), que me parecen necesario de tener en cuenta en lo que se refiere a la economía y el conocimiento en las redes.
- Creo que es necesario mantener un espíritu crítico en el análisis de las redes, en particular de las redes tecnológicas puesto que “si la tecnología no es buena ni mala, tampoco es neutra” (según Kransberg 1985, citado por Castells 1996).

¹⁵ Por ejemplo, la URL del enlace “<http://raw.cs.berkeley.edu/texpoint/index.htm>” es transformada en “<http://raw.cs.berkeley.edu>”.

- En el análisis de la Web, sería un error ignorar los llamados “webometric studies”, y a los cuales no me he referido en este ensayo (Björneborn & Ingwersen, 2004, Ingwersen & Björneborn, 2004, Thelwall & Vaughan, 2004, Thelwall Met al, 2005).
- Pero me parecen mucho más interesante los trabajos que se realizan en el campo del “web mining” (véase Fürnkranz, 2005, para una visión general; como también Oikonomakou & Vazirgiannis, 2005, acerca del “clustering” de documentos de la Web) y ellos con relación a la investigación en curso sobre la Web semántica.

Para concluir diré que estamos atrapados en las redes y que nuestra sola libertad es el conocimiento de ellas. Como dice el verso de Emily Dickinson: “*The web of life is woven*”.

Referencias

- Adamic L. (1999) The Small World Web, *Proceedings of the Third European Conference on Digital Libraries. Lectura Notes in Computer Science* 1696 (Berlin, Springer), p. 443-452
- Adamic L. and Huberman B. (2002) Zipf’s law and the Internet, *Glottometrics*, vol. 3, p. 143-150
- Albert R., Jeong H., and Barabási A.L. (1999). Diameter of the World-Wide Web, *Nature*, vol. 401, p. 130-131.
- Albert R. and Barabási A.L. (2000) Topology of evolving networks: Local events and universality, *Physical Review Letters*, vol. 85, n° 24, p.5234-5237.
- Albert, R. and Barabási, A.L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, vol. 74, n° 1, p. 47-97.
Disponble (24/03/06): http://prola.aps.org/abstract/RMP/v74/i1/p47_1
- Amaral L.A.N., Scala A., Barthélemy M., Stanley H.E. (2000) Classes of small-world networks, *Proc. Nat. Acad. Sci. USA*, vol. 97, n° 21, p. 11149-11152.
Disponble (24/03/06): <http://www.pnas.org/cgi/content/full/97/21/11149>
- Artus P. (2001) *La nouvelle économie*. Paris, Éditions La Découverte (col. Repères)
- Barabási A.L. and Albert R. (1999). Emergence of scaling in random networks, *Science*, vol. 286, p. 509-511.
- Barabási A.L. (2001) The physics of the Web, *Physics World*, 2001, vol. 14, n° 7,
Disponble (15/03/06): <http://www.physicsweb.org/>
- Björneborn L., Ingwersen P. (2004) Toward a basic framework for webometrics, *Journal of the American Society of Information Science and Technology*, vol. 55, n° 14, p. 1216-1227.
- Bolstanki L. et Chiapello E. (1999) *Le nouvel esprit du capitalisme*, Paris, Editions Gallimard.
- Brachman R.J. and Anand T. (1996) The Process of Knowledge Discovery in Databases, in U.M Fayyad, G. Piatetsky-Shapiro, P. Smith, and R. Uthurusamy (editors), *Advances in Knowledge Discovery and Data Mining*. Menlo Park, Calif., AAAI Press & The MIT Press, p. 37-57.
- Bradford S.C. (1934) Sources of information on specific subjects, *Engineering*, vol. 26, p. 85-86
- Brin S. and Page L. (1998) The anatomy of a large scale hypertextual web search engine, *Proceedings 7th WWW*.
Disponble (24/03/06): <http://www-db.stanford.edu/~backrub/google.html>
- Broder, A., R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, J. Wiener. (2000) Graph structure in the Web, *Computer Networks*, vol. 33, p. 309-320.
- Callon M., Courtial J-P., Penan H. (1993) *La scientométrie*, Paris, Presses Universitaires de France, (Que sais-je? n°2727).
- Castells, M. (1996). *The Information Age: Economy, Society, and Culture. Volume 1: The Rise of the Network Society*. Cambridge, MA: Blackwell Publishers.
- Chakrabarti S. Dom B. E. Kumar S. R. Prabhakar R. Rajagopalan S. Tomkins A. Gibson D. (1999) Mining the Web’s Link Structure, *IEEE Computer*, August, p. 60-67.
- Chakrabarti, S. (2000) Data Mining for Hypertext: A Tutorial Survey, *SIGKDD Explorations*, vol. 1, n° 2, p. 1-11.
Disponble (24/03/06): <http://www.acm.org/sigs/sigkdd/explorations/archive.php>
- Chakrabarti S. (2003) *Mining the Web: Discovering knowledge from hypertext data*. Morgan Kaufmann, San Francisco, 2003.
- Courtial J-P., (1990) *Introduction à la Scientométrie*, Paris, Anthropos.
- Cui L. (1999) Rating health Web sites using principles of citation analysis: A bibliometric approach, *Journal of Medical Internet Research*, vol. 1, n° 1.
Disponble <http://www.jmir.org/1999/>

- Degenne A., et Forsé M. (2000) *Les réseaux sociaux*. Paris, Armand Colin. Primera edición 1994.
- Dorogovtsev S.N., Mendes J.F.F., Samukhin A.N. (2000) Structure of growing networks with preferential linking, *Physical Review Letters*, vol. 85, p. 4633-4636.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationships of the Internet topology, *Computer Communication Review*, vol. 29, p. 251-262.
Disponible (24/03/06): <http://www.cs.ucr.edu/~michalis/PAPERS/FFF.pdf>
- Flake G., Lawrence S., Giles C. L., Coetzee F. (2002) Self-Organization and Identification of Web Communities, *IEEE Computer*, vol. 35, n° 3, p. 66-77.
- Foray, D.(2000) *L'économie de la connaissance*, Paris, La Découverte (col. Repères)
- Frank R. and Cook P. (1995) *The Winner-Take-All Society*, Martin Kessler Books at The Free Press.
- Freeman L.C. (1979) Centrality in social networks: I Conceptual clarification, *Social Networks*, vol. 1, p. 215-239.
- Freeman L.C., Borgatti S., White D. (1991) Centrality in valued graphs: A measure of betweenness based on network flow, *Social Networks*, vol. 13, p. 141-154.
- Fürnkranz J. (2005) Web Mining, in O. Maimon and L. Rokach (editors), *The Data Mining and Knowledge Discovery Handbook*, Springer, chapter 42, p. 899-920.
- Garfield. E. (1963). Citation indexes in sociological and historical research, *American Documentation*, vol. 14, p. 289-291.
- Garton, L, C. Haythornthwaite, B. Wellman (1997) Studying Online Social Networks, *Journal of Computer-Mediated Communication*, vol. 3, n° 1
Disponible (24/03/06): <http://jcmc.indiana.edu/vol3/issue1/garton.html>
- Haitun S.D. (1982) Stationary Scientometric Distributions, *Scientometrics*, vol. 4, n° 1, 2, 3.
- Huberman B., Pirolli P., Pitkow J, Lukose R. (1998) Strong Regularities in World Wide Web Surfing, *Science*, vol. 280. n° 5360, p. 95 -97
- Huberman B., Adamic L. (1999) Growth dynamics of the Worl-Wide Web, *Nature*, vol. 401, p. 131
- Huberman B.A. (2001) *The Laws of the Web*. Cambridge, Mass. The MIT Press.
- Ingwersen P. and Björneborn L. (2004) Methodological Issues of Webometric Studies, in Moed H. et al (editors), *Handbook of Quantitative Science and Technology Research*, Kluwer Academic Publishers, chapter 15, p. 339-369.
- Jeong, H., Tomber, B., Albert, R., Oltvai, Z., and Barabási, A.. (2000) The large-scale organization of metabolic networks, *Nature*, vol. 407, p. 651-654.
- Kleinberg, J.M. (1998) Authoritative sources in a hyperlinked environment, *Proceedings of the 9th annual ACM-SIAM Symposium on Discrete Algorithms*, p. 668-677.
Disponible (24/03/06): <http://citeseer.ist.psu.edu/kleinberg99authoritative.html>
- Kleinberg, J.M. (1999) Hubs, Authorities, and Communities, *ACM Computing Surveys*, vol. 31, n° 4 Disponible (27/03/06):
http://www.cs.brown.edu/memex/ACM_HypertextTestbed/papers/10.html
- Kleinberg J. and Lawrence S. (2001) The Structure of the Web, *Science*, vol. 294, p. 1849-1850.
- Kohonen T. (1997) *Self-Organizing Maps*, Berlin, Springer. Primera edición 1995.
- Kohonen T., Kaski S., Lagus K., Salojärvi J., Honkela J., Paatero V., Saarela A. (1999) Self organization of a massive text document collection, in E. Oja and S. Kaski (editors) *Kohonen Maps*, Elsevier, p. 171-182.
- Kosala R. and Blockeel H. (2000) Web Mining Research: A Survey, *SIGKDD Explorations*, vol. 1, n° 2, p. 1-15
Disponible: (24/03/06): <http://www.acm.org/sigs/sigkdd/explorations/archive.php>
- Knoke D. and Kurlinksi J. (1994) *Network Analysis*, SAGE Publications (col. Quantitative Applications in the Social Sciences, n° 28)
- Krapivsky P.L., Redner S., Leyvraz F. (2000) Connectivity of Growing Random Networks, *Physical Review Letters*, vol. 85, p. 4629.
- Kumar, R., Raghavan, P., Rajagopalan, S., and Tomkins, A. (1999) Trawling the Web for emerging cyber-communities, *Computer Networks*, vol. 31, n° 11-16, p.1481-1493
- Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., and Upfal, E. (2000) The Web as a graph, *Proc. 19th ACM SIGACT-SIGMOD-AIGART Symp. Principles of Database Systems*, p. 1-10. ACM Press.
- Kumar R. (2002) The Web and Social Networks, *Computer*, vol. 35, n° 11, p. 32-36.
- Lamirel J-C. (1995) *Application d'une approche symbolico-connexionniste pour la conception d'un système documentaire hautement interactif*, Thèse de l'Université de Nancy 1 Henri Poincaré
- Lotka A.J. (1926), The frequency distribution of scientific productivity, *Journal of the Washington Academy of Sciences*, vol. 16, n° 12, p. 317-23
- Mandelbrot (1954) Structure formelle des textes et communication, *Word*, vol. 10, n° 1, p. 1-27.

- Mandelbrot (1967) Sur l'épistémologie du hasard dans les sciences sociales. Invariance des lois et vérification des prédictions, J. Piaget (ed) *Logique et connaissance scientifique*. Encyclopédie de la Pléiade, Paris, Editions Gallimard, p. 1097-1113.
- Mandelbrot B. (1968) Information theory and Psycholinguistics: A Theory of Word Frequencies, in P.F. Lazarsfeld et al (eds), *Readings in Mathematical Social Science*, Cambridge, Mass., The MIT Press, p. 350-368
- Mandelbrot B. (1975) Arbres de hiérarchie ou de classement et la dimension, capítulo XII de su obra *Les objets Fractals*, Paris, Flammarion, p. 147-152 (citación de la tercera edición de 1989)
- Mercklé P. (2004) *Sociologie des réseaux sociaux*, Paris, La Découverte (col. Repères).
- Merton R. (1973) The Matthew Effect in Science (1968), in *Sociology of science*, capítulo 20, The University of Chicago Press, p. 439-459.
- Milgram S. (1967) The Small World Problem, *Psychology Today*, May 1967, p. 60-67.
- Oikonomakou N. and Vazirgiannis M. (2005) A Review of Web Document Clustering Approaches, in O. Maimon and L. Rokach (editors), *The Data Mining and Knowledge Discovery Handbook*, Springer, chapter 43, p. 921-940
- Pareto (1896) *Cours d'Economie Politique*, Genève, Droz.
- Pennock D. M., Giles C. L., Flake S., Lawrence S., Glover E. (2002) Winners Don't Take All: Characterizing the Competition for Links on the Web, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, n° 8, p. 5207-5211.
- Polanco, X., Boudourides M. A., Besagni D., Roche I. (2001) Clustering and Mapping European University Web Sites Sample for Displaying Associations and Visualizing Networks, in *ETK-NTTS 2001, New Techniques and Technologies for Statistics Exchange of Technology and Know-how, Pre-proceedings of the Conference*, Hersonissos, Crete, 18-22 June 2001, vol. 2, p. 941-944.
- Polanco X. (2002) Clusters, Graphs, and Networks for Analysing Internet-Web Supported Communication within Virtual Community. 7th International ISKO Conference, Granada, Spain, 10-13 July 2002, in *Advances in Knowledge Organization*, vol. 8, p. 364-371.
- Polanco X., Roche I., Besagni D. (2006) User science indicators in the Web context and co-usage analysis, *Scientometrics*, vol. 66, n° 1, p. 171-182.
- Popper K. (1979) *Objective Knowledge: An Evolutionary Approach*. Clarendon Press, Oxford.
- Price D. (1965) Network of Scientific Papers, *Science*, vol. 149, n° 3683, p. 510-515.
- Price D (1976) A General Theory of Bibliometric and other Cumulative Advantage Processes, *Journal of American Society for Information Science*, vol. 27, n° 5, p. 292-306.
- Price D. (1986) *Little Science and Big Science...and Beyond*, Columbia University Press.
- Rousseau R. (1997) Situations: an exploratory study, *Cybermetrics*, vol. 1, n° 1, Paper 1
Disponible: <http://www.cindoc.csic.es/cybermetrics/>
- Ruelle D. (2000) *Hasard et chaos*, Paris, Editions Odile Jacob.
- Scott J. (1991) *Social network analysis: A handbook*. Newbury Park, CA, Sage Publications.
- Simon H.A. (1957) On a Class of Skew Distribution Functions, in *Models of Man: Social and Rational*, chapter 9, New York: John Wiley & Sons, p. 145-164.
- Thelwall M (2003) What is this link doing here? Beginning a fine-grained process of identifying reasons for academic hyperlink creation, *Information Research*, Vol. 8, n° 3. Disponible: <http://informationr.net/ir/8-3/infres83.html>
- Thelwall M., Vaughan L. (2004) Webometrics: An introduction to the special issue *Journal of the American Society of Information Science and Technology*, vol. 55, n° 14, 1213-1215.
- Thelwall M., Vaughan L., Bjorneborn L. (2005) Webometrics, *Annual Review of Information Science and Technology*, vol. 39, p. 81-135.
- Yablonsky A.I. (1985) Stable Non-Gaussian Distributions in Scientometrics, *Scientometrics*, vol. 7, n° 3-6, p. 459-470.
- Wasserman S. and Faust K. (1999) *Social Network Analysis. Methods and Applications*. Cambridge, Cambridge University Press. Primera edición 1994.
- Watts, D.J., and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks, *Nature*, vol. 393 (June 4), pp. 440-442.
- Watts, D.J. (1999). *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton, NJ: Princeton University Press.
- Zipf G.K. (1949) *Human Behaviour and the Principle of Least-Effort*, Addison-Wesley, Cambridge Mass.

Índex

- Atractivo, 10
 - Centralidad, 15-17
 - Grado, 15, 16
 - Intermediación, 15, 16
 - Proximidad, 14, 16
 - Centralización, 17
 - Densidad, 18
 - Efecto de red, 9
 - Enlazado preferencial, 9
 - Grafo, teoría, 11-12
 - Grafos, tipos de, 11
 - Nodos, tipos de, 13
 - Relaciones, tipos de, 12
 - Ley de potencia libre de escala, 6-8
 - Pequeño mundo, 13-15
 - Aparición, 13
 - Definición, 14
 - Propiedades, 14
 - Coefficiente de racimo, 14
 - Largo del camino promedio, 14
 - Red, análisis, 3-30
 - Contenido, 5, 24, 29
 - Estructura, 5, 19, 20, 22-23, 28
 - Uso, 5, 24-25
 - Red tecnológica de la comunicación, 5, 6-10
 - Red social, 3, 24-25
 - Red de conocimiento, 5, 18-19
 - Análisis, modelos, 18-28
 - Citación, de la, 19-20
 - Centros de actividad y autoridades, 19
 - Web comunidades, 19
 - PageRank, 19
 - Co-ocurrencia, de la, 21-25
 - Co-sitios, 22-23
 - Palabras asociadas, 23
 - Co-uso, 24-25
 - Mapas de auto organización, de los, 26-29
 - SOM, 26-27
 - Multi-SOM, 27-28
 - Sistemas no lineales, 6, 8-9
-