



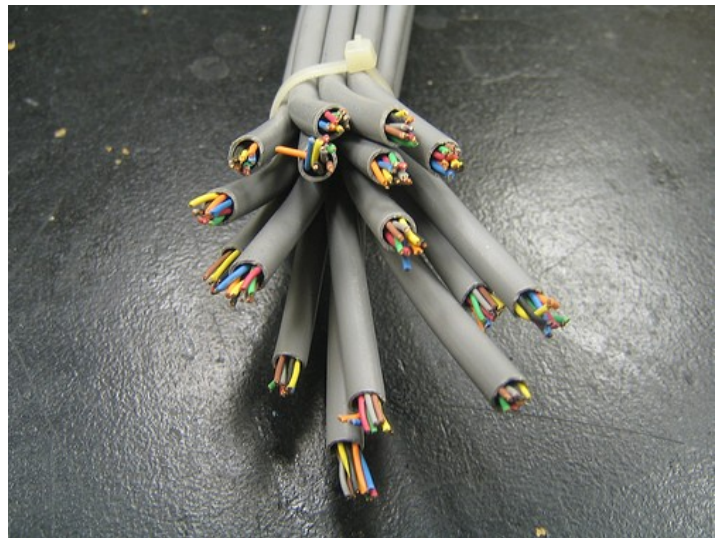
Los centros de gravedad de la Web

Sandra González-Bailón
Oxford Internet Institute, Universidad de Oxford, Inglaterra

Tipo de artículo: Actualidad.
Disciplinas: Sociología, Economía.
Etiquetas: redes sociales, Web, internet, buscadores, información.

El acceso a información en la Web no es independiente de la estructura de red que forman los enlaces entre páginas y dominios. Los buscadores reconstruyen mapas de esas conexiones para identificar los recursos más centrales y darles prioridad en sus algoritmos. Sin embargo, la estructura de la Web no responde sólo a la calidad de los contenidos.

¿Cómo se dibuja el mapa de un mundo cuyos confines se desconocen? La estrategia que siguieron los poderes del mundo en tiempos en los que todavía existía la *terra incognita* fue mandar expediciones que rastrearán caminos y reportaran de vuelta las rutas seguidas. Todavía existen hoy terrenos por descubrir, sólo que son virtuales, y los caminos adoptan la forma de enlaces electrónicos. La Web, a la que la mayor parte de los usuarios accede mediante el uso de buscadores, es una red en constante crecimiento y cambio: cada segundo se añaden y borran incontables páginas, y todos estos cambios tienen lugar de forma descentralizada, sin que haya una autoridad que ponga orden y cense ese flujo constante de información. Mantener un mapa de la Web que, si no completo, esté al menos actualizado es un reto constante para los buscadores, y su estrategia se asemeja mucho a la de los estados de antaño: mandar cartógrafos que recojan suficiente información para reconstruir la topografía que pisaron. La diferencia es que esos cartógrafos adoptan la forma de programas informáticos, o robots, que automatizan la función de seguir enlaces y reportar de vuelta los destinos alcanzados.



(cc) Pascal Charest

Tal y como ocurrió con el descubrimiento del nuevo mundo, esos exploradores electrónicos han permitido dibujar mapas que revelan continentes, y corrientes que revelan la forma en la que navegamos la red. La Web es una red en la que las páginas o documentos publicados son nodos, unidos por enlaces que los conectan. La conexión que une una página a otra no implica que desde la segunda también se pueda llegar a la primera: los enlaces en esta red son como caminos de un solo sentido y a menudo promueven flujos de no retorno. Dada esta característica, la Web está dividida en un centro en el que las páginas están densamente conectadas y una periferia en la que las conexiones son más escasas y dispersas. Estos continentes están identificados en la Figura 1, que reproduce un mapa de la Web basado en las conexiones de más de 200 millones de páginas (Broder et al., 2000; Barabási, 2002: 166-167; Pastor-Satorras y Vespignani, 2004: 143-144). Según este mapa, si iniciáramos un recorrido desde el continente de la izquierda sería fácil llegar al núcleo o continente central, pero no a la inversa: las páginas en esa parte de la Web son esencialmente puntos de partida, no de llegada. Justo al contrario sucede con el continente de la derecha: una vez se llega a él, es difícil encontrar una vía fuera. La mayor parte de las páginas que pueblan el continente de entrada son páginas personales, o dominios recién creados que aún no tuvieron tiempo de llegar al núcleo de la red; las que forman el continente de salida son, de forma significativa, páginas corporativas. El cuarto continente está formado por islas y penínsulas que no pueden alcanzarse desde el continente central: alrededor de un cuarto de todos los documentos están ubicados en esta zona reclusa de la Web.

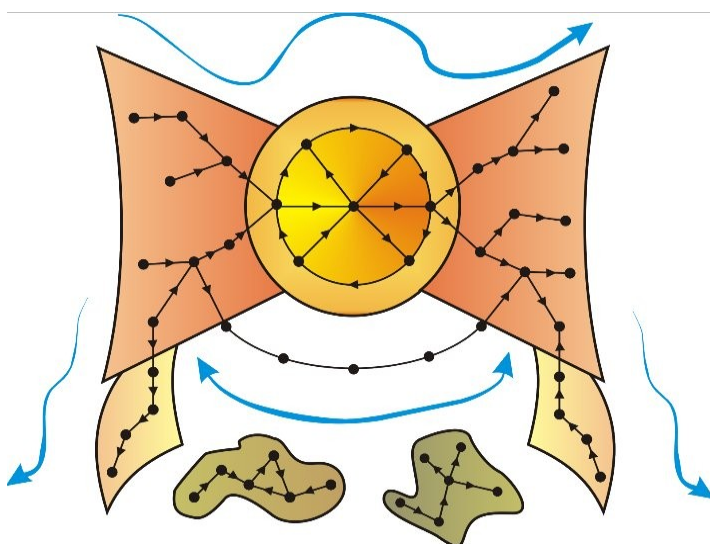


Figura 1.- Los continentes de la Web.

Estos mapas ponen de manifiesto que, como en Roma, casi todos los caminos llegan al mismo destino. Los buscadores han usado esta estructura de red para construir sus algoritmos de organización de los resultados de las búsquedas. Estos algoritmos interpretan los enlaces como votos de confianza: los enlaces que recibe una página son fundamentales para definir no sólo su propia visibilidad, sino también la de las páginas con las que está conectada (Bonacich, 1972; Brin y Page, 1998). Los enlaces enviados por páginas centrales contribuyen mucho más a aumentar la visibilidad de la página de destino que los enlaces enviados por páginas periféricas. Esta distribución de influencia asume que la Web es una red de documentos que funciona de modo

similar a las redes de publicaciones científicas: cuantos más artículos citan una publicación, más valor adquiere ese trabajo, y más valor y visibilidad adquieren los artículos que ese trabajo cita. Sin embargo, un número creciente de investigadores está poniendo en duda la validez de esa metáfora de partida, mostrando que la Web se asemeja más a una red social que a una red de documentos.

Dos estudios recientes analizan la centralidad y audiencia de un millar de dominios en la Web como una función de los recursos y visibilidad de las organizaciones que publican esas páginas (González-Bailón, 2009, en prensa). La Figura 2 muestra la distribución de tres tipos de recursos entre esas organizaciones: centralidad, o número de enlaces que reciben en la Web; tráfico, o número de visitantes que entran en sus páginas; y visibilidad, medida como el número de veces que esas organizaciones son mencionadas por prensa escrita tradicional. Los resultados indican que una minoría de organizaciones acumula la mayoría de los recursos en los tres casos, aunque la desigualdad es mayor en el caso de visibilidad en prensa escrita. La pregunta que los dos estudios plantean es hasta qué punto estas tres distribuciones están relacionadas entre sí. Según los análisis, las organizaciones más ricas y más visibles en medios de comunicación tradicional también son las más centrales y las más visitadas en internet.

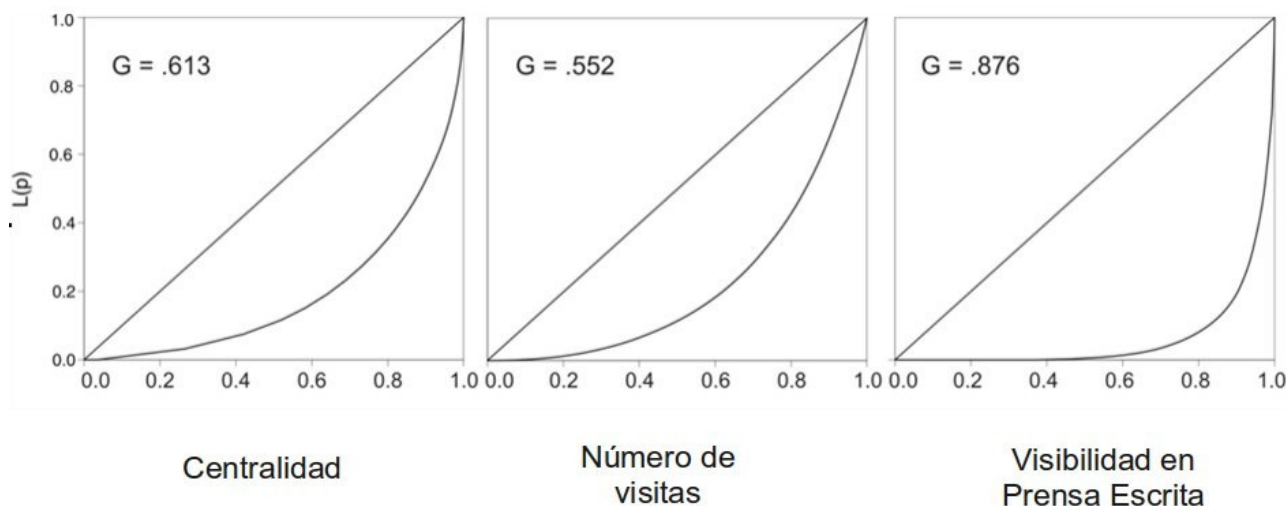


Figura 2.- Desigualdad en la distribución de recursos de los sitios Web analizados, medida según el Coeficiente de Gini (véase http://es.wikipedia.org/wiki/Coeficiente_de_Gini_para_una_definición). A mayor coeficiente, mayor desigualdad. Un coeficiente de 1 indica que una organización acumula todos los recursos (vínculos, visitantes y menciones), mientras que un coeficiente de 0 indica que todas las organizaciones acumulan el mismo número de recursos.

Estos análisis, y la literatura en la que se insertan, ponen de manifiesto que los enlaces esconden estrategias y alianzas que denotan asimetrías de poder. Al fin y al cabo, si todos los caminos llegaban a Roma era porque Roma era la capital del Imperio. Las características de los que publican contenidos en la Web son importantes para entender por qué algunas fuentes de información son más visibles y tienen más impacto que otras; y esto, a su vez, genera consecuencias sobre la pluralidad y diversidad de la información disponible. En su crecimiento, la Web seguirá siendo descentralizada y difícil de controlar por una sola autoridad, pero los buscadores, en su función de guardianes de esa información, están fortaleciendo ciertos centros de gravedad de los que, como mínimo, merece la pena saber más.

Referencias

Broder, A., Kumar, R., Maghoul, F., Raghavan, P. y Rajagopalan, S. (2000) Graph structure in the web. *Computer Networks*, 33, 309-320.

Barabási, A. L. (2002) *Linked. The New Science of Networks*. Cambridge, MA: Perseus.

Pastor-Satorras, R. y Vespignani, A. (2004) *Evolution and Structure of the Internet. A Statistical Physics Approach*. Cambridge: Cambridge University Press.

Bonacich, P. (1972) Factoring and weighting approaches to clique identification. *Journal of Mathematical Sociology*, 2, 113-120.

Brin, S. y Page, L. (1998) The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30, 107-117.

González Bailón, S. (2009) Opening the black box of link formation: Social factors underlying the structure of the web. *Social Networks*, 31, 271-280.

González Bailón, S. (en prensa) Traps on the web: The impact of economic resources and traditional news media on online traffic flow. *Information, Communication & Society*.

Manuscrito recibido el 20 de octubre de 2009.
Aceptado el 22 de noviembre de 2009.

