Algunas medidas y relaciones en la blogosfera hispana

POR **JUAN JULIÁN MERELO Y FERNANDO TRICAS**

El estudio cuantitativo del tamaño y de las relaciones propias de la *blogosfera* hispana nos ayuda a comprender mejor el fenómeno *blog* en el ámbito de las lenguas hispanas, hasta ahora no muy bien representado en los estudios internacionales, centrados habitualmente en la *blogosfera* angloparlante. Una de las principales conclusiones es que su tamaño (medido en número de bitácoras) ha crecido de forma exponencial.

A los efectos del estudio cuyos resultados presentamos, se puede considerar la blogosfera como el conjunto de los *blogs* o bitácoras (1) que, en su lado humano, está representada por la comunidad formada por sus autores, que, individual o colectivamente, comparten información y opiniones en sus sitios. Evidentemente, si miramos al fenómeno *blog* en su generalidad, la blogosfera tiene otros componentes humanos: los que comentan y los lectores habituales y/o casuales, por ejemplo. Sin embargo, aunque estos últimos influyen en la dinámica de la misma, es mucho más interesante mirar sólo a las bitácoras en sí y los enlaces que se producen entre ellas, porque, primero, es una información medible e inequívoca (la dirección de un sitio o página web es única) y, segundo, plasma una relación que se podría expresar como *conoce-a*. El tratar la blogosfera como el conjunto de todos los lectores y escritores de *blogs*, aunque se ha abordado en otros trabajos (Ferreras, 2005) requiere quizás de un enfoque diferente, no reticular, que queda fuera de los límites de este trabajo.

Precisamente, lo que intentaremos hacer es acercarnos a la blogosfera hispana desde la teoría de redes sociales complejas, practicando una estimación de su tamaño y de su configuración global. Una de estas características globales interesantes es el hecho de que se trate de un mundo pequeño o no (Watts, Strogatz, 1999), es decir, que la distancia media entre dos puntos cualesquiera de la blogosfera sea pequeña, y que el aumento de esta distancia, al añadir nuevos nodos, sea lento (2). El hecho de que una red sea un ?mundo-pañuelo? (traducción

Fundación Telefónica

más exacta del término inglés *small-world*, mundo pequeño) implica una transferencia de información rápida entre los diferentes nodos, lo que a su vez comporta también cambios de estado rápidos.

Por otra parte, una característica interesante en bastantes redes de origen social o biológico es que sean ?libres de escala? (3) (Albert, Barabási, 2002); es decir, que no haya ningún número de enlaces *preferido* o *común*. En las redes libres de escala, el número de enlaces se organiza según una ley de potencias: hay muchos nodos con pocos enlaces, pero existe la posibilidad de que haya algunos nodos con un número de enlaces uno o varios órdenes de magnitud mayor que la mayoría de los demás. A esto alude el fenómeno de las ?colas largas? o *long-tail*: en algunos casos, el volumen total de enlaces de los nodos con menos enlaces puede ser superior que el de los sitios con muchos.

Volvamos entonces a la red social de la que se ocupa este artículo: la blogosfera hispana. Las primeras medidas sobre su crecimiento muestran que el aumento de tamaño en los últimos meses ha tenido un ritmo similar al de la blogosfera global (Sifry, 2005): se ha multiplicado por veinte desde los primeros estudios que se hicieron hace unos dos años (Tricas, Ruiz, Merelo, 2003; Merelo, Tricas, Ruiz, 2003), aunque, en realidad, la mejor estimación que se puede hacer sobre el tamaño total es una suposición informada. Es muy posible, sin embargo, que el número de bitácoras desconocidas haya crecido de la misma forma que el de las conocidas.

El tamaño alcanzado complica que se puedan hacer cálculos sobre el grafo de hiperenlaces de la blogosfera; por eso, en este trabajo pretendemos buscar estrategias alternativas para medir el mismo tipo de estructuras en una selección de las bitácoras más representativas, para saber si somos capaces de establecer alguna clase de relación entre estos resultados y aquéllos.

Para empezar, hay que establecer algún criterio acerca de qué se considera ?representativo? dentro de la blogosfera hispana. Sobre esta muestra representativa, se establecerá el grafo de conexiones de la misma y se hará una serie de medidas que nos permitirán identificar subestructuras y calcular cuáles son las bitácoras más notables por su influencia o su posición dentro de la red. Dentro de la blogosfera hispana, y dado que la localización geográfica es, en la mayoría de los casos, difícil de determinar, optamos en su día por incluir a todas aquellas que estén escritas en alguno de los idiomas oficiales de España, lo que incluye también a muchas bitácoras localizadas en países latinoamericanos (con la excepción de Brasil, cuando somos capaces de establecerlo claramente).

Nuestra experiencia en el estudio de la blogosfera viene dada por el desarrollo de herramientas alrededor de los *weblogs*. En particular, el *blogómetro* (4), que recorre diariamente la blogosfera extrayendo los enlaces nuevos que aparecen en las bitácoras y almacenándolos en una base de datos. También se ha hecho una serie de estudios, más exhaustivos, sobre Blogalia (5) (Merelo, Prieto, Tricas, 2004), una comunidad de la cual se tienen datos desde su creación, y que presenta una dinámica que es representativa de la que sigue el resto de la blogosfera.

Tamaño

Fundación Telefónica
Telefónica

Llevamos recopilando información desde finales del año 2002, cuando, por la poca representación de la blogosfera hispana en los servicios de publicación y directorios más populares (6), y por la imposibilidad de separar ?nuestra? blogosfera del resto, decidimos crear nuestra propia herramienta, para tratar de obtener datos fiables y, de paso, ayudar a los integrantes de la propia blogosfera a tener acceso a dicho tipo de servicios, con datos más próximos. La idea era publicar un *ranking* diario de los sitios mas enlazados desde las bitácoras de nuestra lista.

Nuestra herramienta se desarrolla de manera abierta y el programa que se ocupa de la recogida de datos, así como la lista de bitácoras, están disponibles en *SourceForge* (7). Aunque contamos con la colaboración habitual de algunos de los proveedores más importantes (Blogia, Acelblog, Ya.com) para actualizar nuestros datos, otros no proporcionan información acerca de bitácoras alojadas por idioma, por situación geográfica o ninguna en absoluto. También algunos propietarios de bitácoras interesados en el proyecto nos facilitan los datos de sus propios sitios.

En el momento de escribir estas líneas, contamos con 102.498 bitácoras en nuestro listado, que se distribuyen por alojamientos según se recoge en la tabla 1 (1) (sólo aparecen alojamientos con más de 1.000 bitácoras):

En esta lista están mal representados (a la baja) los MSN Spaces (dominio *msn.com*) de Microsoft (por su reciente aparición) y las bitácoras albergadas en Blogger (dominio *blogspot.com*), por la dificultad que entraña encontrar nuevas bitácoras publicadas en este sitio, que es el más popular y uno de los más longevos.

El crecimiento de la blogosfera sigue una tendencia claramente exponencial, duplicándose cada 5-6 meses, como puede apreciarse en la figura 1 (2).

Relaciones

Una de las primeras preguntas que se puede plantear es si este crecimiento sigue las reglas del ?enlazado preferencial? (Albert, Barabási, 2002), es decir, si las bitácoras existentes enlazan, con preferencia, a aquellas que ya tienen enlaces; o dicho de otra forma, la probabilidad que tiene un nodo de recibir un nuevo enlace crece con el número de enlaces existentes. Esta dinámica daría lugar a una red que sigue una ley de potencias (Shirky, 2003; Kottke, 2003) en una serie de parámetros; por ejemplo, en un gráfico que represente el número de enlaces vs. orden. Y efectivamente, así sucede, como se muestra en la figura 2 (3). Este gráfico ha sido obtenido a partir de la lista de los sitios más enlazados, confeccionada a partir de los datos del blogómetro y aumentada con los sitios más visitados entre los usuarios de la blogosfera hispana del sistema de estadísticas Nedstat (8); de allí se ha extraído una lista de aproximadamente 500 sitios, y se ha buscado en Technorati el número de enlaces entrantes (9).

La ley de potencias ajustada a las 100 primeras bitácoras es f(x)=1192.61x-0.475358, donde *x*

Fundación Telefónica
Telefónica

es la posición que ocupa la bitácora, puestas por orden decreciente de número de enlaces entrantes. Esta ley de potencias es interesante: como el exponente es menor que uno, daría lugar a la clásica ?cola larga? (Shirky, 2003), es decir, la suma de los enlaces de los sitios menos enlazados sería mayor que la de los enlaces a los sitios más enlazados; la multitud de sitios pequeños de la blogosfera superaría en importancia a los sitios grandes. En todo caso, debido al sesgo en el muestreo (sólo se han tomado 500 bitácoras), el ajuste más adecuado en este caso sería una ley de potencias con corte exponencial (10); aun así, se puede deducir que el hecho de que una bitácora sea más enlazada que otra no induce un sesgo demasiado fuerte en la creación de nuevos enlaces, y que, fuera de las bitácoras más enlazadas, el número de enlaces (que discutiblemente podríamos asociar a popularidad) disminuye rápidamente.

La segunda característica macroscópica que se puede estudiar es el hecho de que se trate de una red *small-world* (pequeño mundo) o no. Para ello se han tomado, usando también el API de Technorati, todos los enlaces entrantes a las bitácoras seleccionadas. A partir de ahí, se han dejado sólo los enlaces *entre* ese grupo. De los 42.700 enlaces iniciales, el conjunto final contiene sólo alrededor de 3.600, es decir, que este grupo ?elitista? inicial no es endogámico en el sentido de que el número de enlaces procedentes de fuera es mayor que el de los enlaces del propio grupo (que representan menos de un 10 por ciento del total). Esto ha supuesto una segunda selección: algunos de los nodos del grupo inicial no tienen ningún enlace dentro de él. El nuevo grupo sólo tiene 431 componentes; de esos, 428 están conectados entre sí, formando otros tres un pequeño grupo aparte. Curiosamente, en un estudio hecho a principios de 2004 (Merelo, 2004), había dos grandes islas bien definidas que, grosso modo, correspondían a las bitácoras españolas y las iberoamericanas.

El hecho de que se trate de una red ?mundo pequeño? no está sólo relacionado con el diámetro de la red, que es 3,402, sino con el coeficiente de *clustering*. El coeficiente de *clustering* mide la transitividad de las relaciones, es decir, para cada 3 nodos, en qué proporción sucede que cuando A está relacionado con B, B está también relacionado con C. En este caso, es 0,236, mucho mayor que el correspondiente a una red aleatoria *Erdös-Renyi* (11) con el mismo número de nodos y grado de conexión medio, que sería 0,034. La blogosfera hispana es, por lo tanto, un pañuelo ? al menos a este nivel ? con la apariencia que se muestra en la figura 3 (4).

Este gráfico tiene el aspecto general de los grafos que siguen una ley de potencias: un núcleo de sitios bastante interconectados, con una serie de ?tentáculos? que se relacionan con esta zona central y que, a su vez, forman pequeños grupos con vínculos fuertes entre sí. En todo caso, es curioso observar que hay, si no transferencia ideológica, sí conocimiento y enlace mutuo entre todas las bitácoras que forman esta representación de la blogosfera. Esto parece establecer una diferencia con la blogosfera global donde las bitácoras posicionadas ideológicamente aparecen casi aisladas, desde el punto de vista de los hiperenlaces (Adamic, Glance 2005).

Conclusiones

En este artículo hemos presentado algunos datos cuantitativos sobre la blogosfera hispana; se

Fundación Telefónica
Telefónica

ha puesto de manifiesto su crecimiento en los últimos dos años, y las cualidades de la red que forman las bitácoras: es un mundo pequeño que sigue una ley de potencias; además, su cohesión, es decir, la conexión entre sus diferentes grupos ha aumentado desde los últimos estudios realizados a principios del año pasado.

De la misma forma, el dinamismo de la blogosfera se ha puesto de manifiesto por el pequeño exponente de la ley de potencias ajustada, y su grado de apertura por el hecho de que incluso bitácoras muy diferentes ideológicamente o temáticamente están conectadas directamente, o por una media de aproximadamente tres grados de separación.

En resumen, consideramos útil mantener un observatorio de la blogosfera hispana, para comprobar el efecto que tienen fenómenos tales como la introducción de millones de bitácoras, como lo sucedido recientemente con *MSN Spaces*, eventos como el 11M y su efecto en la creación de nuevas agrupaciones en la estructura reticular de la blogosfera. Asimismo, resultaría interesante correlacionar todos estos fenómenos con la creación de opinión pública y la propagación de memes en la blogosfera.

Bibliografía

ADAMIC, L. GLANCE, N.: «The Political Blogosphere and the 2004 U.S. Election: Divided They Blog? , *2nd Annual Workshop on the Weblogging Ecosystem: Aggregation, Analysis and Dynamics*, WWW2005, Japan, http://www.hpl.hp.com/research/idl/papers/politicalblogs/

ALBERT, R., BARABASI, A.L.: «Statistical Mechanics of Complex Networks», *Reviews of Modern Physics*, núm. 47, 2002, http://arxiv.org/abs/cond-mat/0106096

FERRERAS, 2005: «II Encuesta: los *weblogs* más leídos», publicado en http://tintachina.com/archivo/ii_encuesta_los_weblogs_mas_leidos.php (último acceso 1 de agosto de 2005).

FUMERO, A.: «El abecé del universo *blog*», *TELOS,* 65, octubre-diciembre, 2005.

KOTTKE, J.: *Screw the Power law, Embrace the Power Law*, http://www.kottke.org/03/02/screw-power-law

MERELO, J. J., TRICAS, F., RUIZ, V.: «Do we live in a Small World? Measuring the Spanish-speaking blogosphere», *Proceedings of BlogTalk, An European Conference on Weblogs,* Viena, Austria, May 23-24, 2003, http://geneura.ugr.es/~jmerelo/habilitacion2005/papers/65.pdf

MERELO, J. J.: *Mapas de la* blogosfera, http://atalaya.blogalia.com/historias/15216

MERELO, J. J., TRICAS, F., RUIZ, V., ORIHUELA, J. L.: «*Revisiting the Spanish blogosphere*»,

Fundación Telefónica
Telefónica

Blog Talks 2.0, Viena, Austria, 2004,
http://webdiis.unizar.es/~ftricas/Articulos/Revisiting%20the%20Spanish%20blogosphere.pdf

MERELO, J. J.; PRIETO, B.; TRICAS, F.: *Blogosphere community formation, structure and visualization*. Actas Blogtalk 2.0, Viena, Austria, Julio 2004. Notas en directo de la charla en http://joi.ito.com/joiwiki/BlogTalkNotesPanel3. Disponible también en http://webdiis.unizar.es/~ftricas/Articulos/Blogosphere%20community%20formation.pdf

SHIRKY, C.: *Power Laws, Weblogs, and Inequality*, http://www.shirky.com/writings/powerlaw_weblog.html. First published February 8, 2003 on the ?Networks, Economics, and Culture? mailing list. Accedido por última vez el 2 de agosto de 2005.

SIFRY, D.: *State of The Blogosphere, March 2005, Part 1: Growth of Blogs,* http://www.sifry.com/alerts/archives/000298.html

WATTS, D. J.; STROGATZ, D. J.: *Collective Dynamics of ´Small-World´ Networks*, Nature, 393:440-442 (1998).

Fundación Telefónica
Telefónica