

Análisis de Redes de Influencia en Twitter

Carlos Guadián Orta¹, Francisco Manuel Rangel Pardo², Javier Llinares Salas³

¹ Director de Comunicación e Innovación, Autoritas Consulting, SA
carlos.guadian@autoritas.es

² Director de Tecnología, Autoritas Consulting, SA
francisco.rangel@autoritas.es

³ Director General, Autoritas Consulting, SA
javier.llinares@autoritas.es

Abstract: La información es poder sólo cuando la información es compartida y produce conocimiento colectivo susceptible de mejorar cualquier aspecto de la sociedad. En este sentido las redes sociales (telemáticas) están jugando un papel revolucionario pero dónde perderse en el anonimato es cada vez más fácil. Si tenemos un mensaje y queremos hacerlo llegar a un gran número de personas debemos saber a través de quién debemos transmitirlo. El presente proyecto se marca como objetivo determinar las redes de influencia y los influenciadores principales para un tema de interés a partir del análisis de redes sociales y sus métricas, presentando como ejemplo los resultados del análisis “Topología de la Comunidad Política Española en Twitter”

Keywords: Twitter, Análisis de Red Social, Red de Influencia, Interacción de Usuarios, Modelado de Red Social, Análisis de Usuarios

1 Introducción

Twitter es una red social de mensajes cortos que ha revolucionado el modo de comunicarse de las personas, y por lo tanto de conformar redes de influencia en torno a los temas que más interesan a sus usuarios.

La red social twitter se constituye por medio de lo que se denominan seguidores y seguidos. Un usuario sigue a los usuarios que considera de su interés y es seguido por usuarios que le consideran a él por su parte también de interés. Intrínsecamente estas redes se verán conformadas por usuarios de intereses similares, bien sean políticos, económicos, de ocio, sanidad... lo que denominamos temas o grupos semánticos primarios.

La información en twitter fluye de manera viral a partir de la creación de un mensaje y su posterior redifusión por la red de influencia de la persona que lo emite, y de las redes de los que lo reciben y deciden reemitirlo, y así sucesivamente hasta que el mensaje muere.

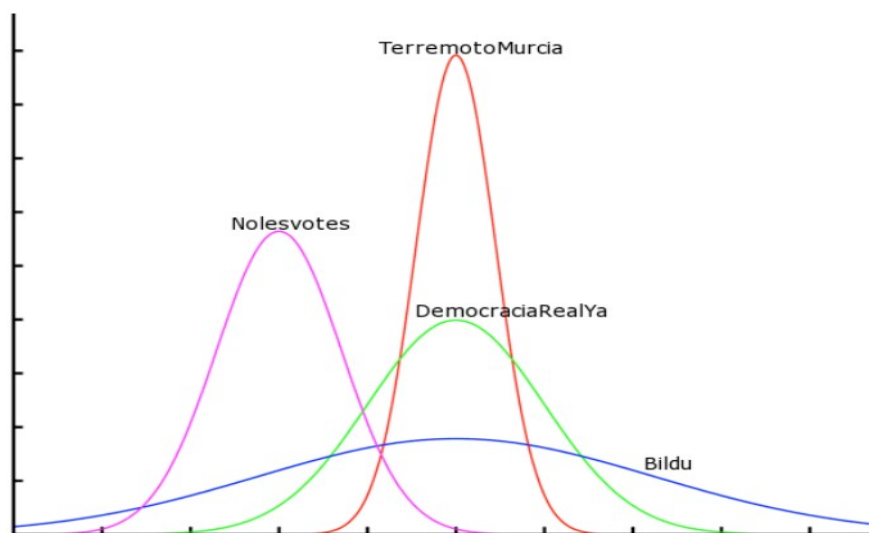


Fig 1. Ciclo de vida de algunos mensajes (simulación informática normalizada no real)

La definición clásica de influencia, por ejemplo tomada de la tercera acepción de la RAE (<http://buscon.rae.es/draeI/SrvltConsulta?LEMA=influencia>) es la siguiente:

“Persona con poder o autoridad con cuya intervención se puede obtener una ventaja, favor o beneficio”

En comunicación, y en este caso en comunicación online, la influencia se traduce en la capacidad de hacer llegar el mensaje al mayor número posible de individuos, lo que clásicamente en prensa o televisión sería el número de impresiones, y que es objetivo principal de la política de promoción del marketing mix de cualquier entidad, sea empresa, partido político, administración pública, etcétera.

El objetivo de la investigación es determinar de manera cuantitativa, en base a métricas del análisis de redes sociales [14, 15, 16], los usuarios más influyentes para un determinado tema o grupo semántico como base para la recomendación de los círculos en los que se debe mover una determinada entidad si quiere hacer efectivo su mensaje.

La primera cuestión que aparece es por tanto la siguiente, *¿cómo podemos construir un grafo social de twitter para determinar quién tiene influencia en un tema que nos interesa?*

Pues bien, este problema parte de una división en dos problemas diferenciados y que dan lugar a sendas investigaciones paralelas:

- La determinación de los **temas** sobre los que los usuarios tienen interés. Se trata de dividir a la población de manera categorial en aquellos temas sobre los que tiene interés. Este análisis no se puede abordar de la manera clásica

(encuestas, muestras...) principalmente por el tamaño de la población (cientos de miles para una pequeña comunidad autónoma, millones para un país) y porque existen metodologías automáticas que permiten una aproximación bastante acertada, estas son, a partir de la recuperación de información, las técnicas de cluster y/o clasificación de procesamiento de lenguaje natural y aprendizaje automático. Esta investigación queda al margen de la que aquí se presenta pero es la base necesaria para la consecución del objetivo, por lo que es necesario conocer y de esta premisa se parte de que los usuarios son etiquetados, bien por lo que hablan, bien por lo que en su bio dicen que les interesa, en los diversos temas o grupos semánticos de primer nivel (pe. sanidad, política, tecnología...)

- La determinación de las **redes sociales** que se conforman y su análisis e interpretación. Partiendo de la población interesada en un determinado tema, muestra categorial de la población completa, aparece la cuestión de *¿cómo agrupamos a esa población y la relacionamos para construir una red susceptible de análisis estructural?*, *¿qué métricas utilizamos para identificar influenciadores?*, *¿cómo comparamos los resultados de diferentes experimentos?*, *¿cómo identificamos al influenciador o grupo de influenciadores clave para lanzar un mensaje?*, *¿cómo extrapolamos los resultados para predecir el impacto de un mensaje?*, en resumen, *¿cómo modelamos las relaciones y la interacción entre usuarios?*

Todas las cuestiones relacionadas con la investigación propuesta buscan responder a la siguiente pregunta:

Si estoy interesado en transmitir un mensaje con éxito, ¿a quién me tengo que “arremar”?

El presente trabajo enfoca el análisis de red de influencia en la obtención de la “Topología de la Comunidad Política Española en Twitter” tras las elecciones del 20N. A continuación se presenta el marco teórico sobre el que se sustenta la investigación, dando paso a la metodología utilizada para llevarla a cabo. Se identifica el dataset sobre el que se trabaja y se presentan los resultados experimentales, comentando los mismos desde un enfoque estructural. Por último se presentan las conclusiones y las propuestas de trabajo futuro para mejorar la construcción de modelos de análisis de influencia.

2 Marco teórico de referencia

La antropología, sociología y otras ciencias sociales incorporan el análisis estructural en sus investigaciones con la ayuda de físicos y matemáticos, quienes traen la teoría de grafos como estandarte, tal y como [8] esquematiza en la siguiente imagen, naciendo de este modo el ARS (Análisis de Redes Sociales).

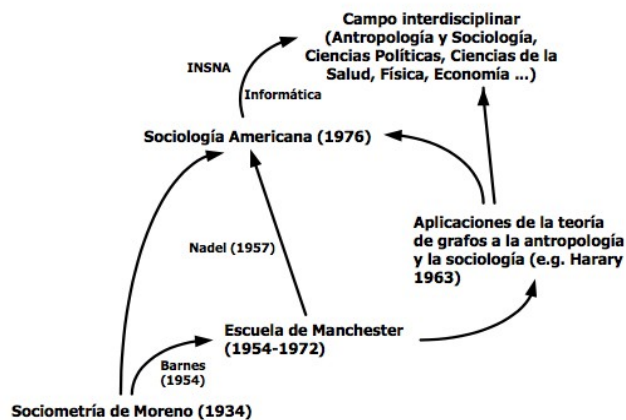


Fig 2. El desarrollo del análisis de redes sociales

El ARS se basa en el estudio de modelos relacionales, “lo que implica en la práctica la imposibilidad de realizar algún tipo de muestreo, ya que es necesario para el análisis contar con información de todos los actores y todos los lazos que componen la red social “ [11], aunque por ejemplo [9] propone un tipo de muestreo basado en redes parciales.

El quid está entonces en qué criterio de muestreo utilizar, o en cómo delimitar la población que conformará la red social de interesados, para lo que en 1983 [12] definieron una tipología de criterios para la delimitación de la población basada en el cruce de dos perspectivas metateóricas, la realista y la nominalista, y de tres focos definicionales, los actores, las relaciones y las actividades, y una combinación de los mismos, resumiéndose en la siguiente tabla:

Perspectiva metateórica	Foco definicional			
	Atributo de los actores	Relación	Participación en actividades	Combinación de focos
Realista	I	III	V	VII
Nominalista	II	IV	VI	VIII

Fig 3. Tipología de delimitación poblacional

Y una vez determinada la población de estudio por uno de los tipos anteriores, se utiliza la teoría de grafos para analizar la red social así modelada.

La teoría de grafos, inicialmente propuesta al menos parcialmente por [4] para la resolución del problema de los puentes de Königsberg, tienen amplia aplicación en campos como la biología, por ejemplo representando habitats naturales y senderos de

migración de los animales, en diseño de circuitos secuenciales, modelados como autómatas de estados finitos, en comunicaciones, para determinar rutas óptimas de transporte o en telecomunicaciones, para determinar rutas óptimas de transporte de paquetes de datos, o incluso en gestión de proyectos (grafos PERT) dónde las tareas se modelizan como nodos y los tiempos entre tareas y su secuencia como vértices y se utiliza la teoría de grafos para analizar caminos críticos.

Pero ni la teoría de grafos, ni los enfoques del ARS pueden ser utilizados, al menos directamente, en los modelos de comunicación de la mass media, con la televisión a la cabeza, debido a que la comunicación es totalmente direccional, es decir, el medio emite y el usuario recibe la información, sin retorno, sin feedback.

Se inventan por tanto medios para determinar los posibles impactos que recibe un determinado mensaje a través de un canal estimando de este modo la influencia del mismo, naciendo índices como el OJD (<http://www.ojdinteractiva.es/ultimos-medios-auditados.php>).

Con el nacimiento de la web son otra serie de campos de investigación nuevos en los que surge el interés del análisis de la influencia, los relacionados con los motores de búsqueda y la recuperación de información. La web nace como un conjunto cada vez mayor, con un crecimiento exponencial de páginas web que comunican también en un sentido pero que tienen cierto feedback, y es la petición a discreción del usuario, al que había que darle los resultados más relevantes según su interés.

Nacieron algoritmos que pretendían ordenar los documentos no sólo por la adecuación de su contenido a lo solicitado por el usuario, sino que estos además fueran importantes. A partir de la observación de los índices bibliográficos basados en la coreferencia de los autores de artículos científicos nace HITS, propuesto por [5, 6, 7], y que utiliza los conceptos de authority y hub bien conocidos en ARS. Poco después, [1] diseñan PageRANK, un algoritmo basado en conceptos similares a HITS que puntúa tanto mejor un documento cuanto mejor ranking tenga, y este ranking se calcula de manera recursiva a partir de un análisis de la red de enlaces entre documentos, otorgando finalmente un número a cada uno de ellos que permite ordenarlos en función de su importancia ó influencia.

Las redes sociales como Twitter parten de todas las ideas anteriores, principalmente las de los motores de búsqueda, pero arrastran influencias de los medios de comunicación tradicionales y así se habla de ciertos indicadores, estándares de facto en la actualidad, como:

- Seguidores: Es el número total de seguidores del usuario, lo que da una estimación de los que directamente visualizarán un mensaje que el usuario emita y por lo tanto una primera medida directa de su influencia, aunque no toma en cuenta más características de la subred del usuario
- Influencia de red: Este valor no sólo tiene en cuenta tus seguidores sino su influencia, para lo que de manera recursiva se calcula el número de seguidores que tienen y la influencia de estos
- Amplificación de red: Es un valor que muestra la influencia directa que tiene un usuario sobre su red, es decir, cuando emite un mensaje, cuál es la probabilidad de que ese mensaje sea retransmitido y por lo tanto amplificado.

- Alcance real: Es un indicador que a partir de la influencia de red y la amplificación de red calcula el valor de audiencia real a la que puede llegar el usuario con un mensaje. Es lo más parecido al número de impresiones medidos en los mass media, pero calculados de manera cuantitativa a partir del análisis de la red
- Otras medidas relacionadas con la pasión, la fuerza, el alcance de los mensajes emitidos por un usuario

En todos estos casos se parte de un análisis de red social creada a partir de los seguidores y los seguidos, y de los mensajes que se reenvían de un usuario, pero todos ellos son análisis centrados en la influencia global de un usuario y no tienen en cuenta el resto de métricas que el ARS puede extraer de una determinada red, y utilizar estas métricas para sugerir influenciadores a partir de un determinado tema.

3 Metodología de Investigación

Se ha pretendido hacer un análisis complementario y desde otro enfoque metodológico al cuantitativo y/o categorial realizado más comunmente. Por ello el estudio no ha tenido en cuenta información como:

- Listas en las que aparecen
- Número de tweets publicados

Por contra se ha centrado el estudio en un enfoque estructural, determinando a “qué usuarios” le dan importancia los partidos/candidatos en Twitter en las elecciones generales del 20 de noviembre, a qué usuarios se les presta más atención y cómo se relacionan entre ellos constituyendo comunidades, y quiénes son los usuarios clave de esas comunidades.

Dada la lista de candidatos/partidos políticos descrita en el siguiente apartado, se ha obtenido todos sus amigos (follow recíproco) y las relaciones existentes entre estos usuarios, y se ha realizado mediante la herramienta Gephi (gephi.org) un análisis de la red social para determinar cuáles son los usuarios más relevantes debido a su posición dentro de la estructura de la comunidad y cuál es el número de sus conexiones.

Las consideraciones metodológicas tenidas en cuenta son las siguientes:

- Para realizar al análisis y la representación se ha considerado que es un grafo dirigido
- Los nodos son los usuarios que conforman el grupo de amigos
- El número de nodos representados en el grafo no corresponde con el total de amigos de cada candidato/partido. Aquellos que no siguen a nadie en la comunidad, excepto el candidato, se han desestimado ya que son nodos aislados
- El tamaño de los nodos está de acuerdo con el grado de centralidad eigenvector (http://www.faculty.ucr.edu/~hanneman/nettext/C10_Centrality.html#Eigenvector)

- Los colores de los nodos indican comunidades, para poder definirlos se ha hecho un cálculo de modularidad con el algoritmo de detección de comunidades de Vincent D. Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre descrito en [13]
- Su posición (la del nodo) en el grafo depende de su gravedad respecto a otros nodos. Cuanto mayor foco de atracción tenga (más followers) más estará en el centro de los que gravitan a su alrededor.
- Las aristas (líneas que unen los puntos) indican una relación
- El color de las aristas es el mismo que el nodo de origen. Es decir, tiene el color del usuario que sigue al otro.

Para completar el análisis se han superpuesto las diferentes comunidades con el objetivo de ver como gravitan unas con otras en función de sus usuarios comunes. Cuantos más lazos en común tengan dos comunidades más cercanía tendrán. Este ejercicio coloca a las diferentes comunidades en posiciones de afinidad por la cantidad de relaciones comunes.

Para poder ver la estructura de las diferentes comunidades y sus tamaños, se deberá tener en cuenta que cuanto más compacta sea una comunidad indicará más número de relaciones existentes entre sus miembros. En este caso los usuarios relevantes de estas comunidades seguramente conseguirán mucho más feedback para sus publicaciones, ya que su comunidad actuará como una caja de resonancia.

Por otro lado aquellas que, aunque con más número, tienen menos densidad de relaciones entre sus miembros, generarán menos respuesta que las más densas.

Lo que nos acaba mostrando el estudio, representado finalmente por el grafo, es el posicionamiento ideológico, por proximidad, de la comunidad política española en Twitter.

4 Datos experimentales

La recogida de datos se ha realizado con la herramienta Cosmos para la Escucha Activa de Autoritas Consulting, recopilando y procesando la información emitida en Twitter durante la campaña de elecciones de 2011, recopilando un total de 941.801 tuits emitidos por 207.867 usuarios únicos.

Para el análisis de red, de la anterior cifra se han tenido en consideración las cuentas que están utilizando en público los diferentes partidos para hacer su seguimiento de la campaña electoral en Twitter, estos son:

@ConRubalcaba (PSOE)	@MarianoRajoy (PP)	@Cayo_Lara (IU)
@CiuDuran2011 (CiU)	@Jerkoreka (PNV)	@AlfredBosch (ERC)
@UPyD (UPyD)	@Juralde (equo)	@ForoAsturias (Foro Asturias)
@AmaiurInfo (Amaiur)		

Tabla 1. Cuentas públicas de candidatos y partidos
Obteniendo un total de 10.338 usuarios únicos (nodos) en la red, relacionados entre sí en un total de 171.532 relaciones (aristas), conformando un diámetro de red de 12 saltos. Los datos se totalizan en la siguiente tabla:

Tuits	Usuarios	Nodos	Aristas	Diámetro
941.80	207.87	10.34	171.53	12

Tabla 2. Dastaset de experimentación

5 Resultados experimentales

A continuación se presenta el grafo resultante del ARS construido según la metodología presentada.

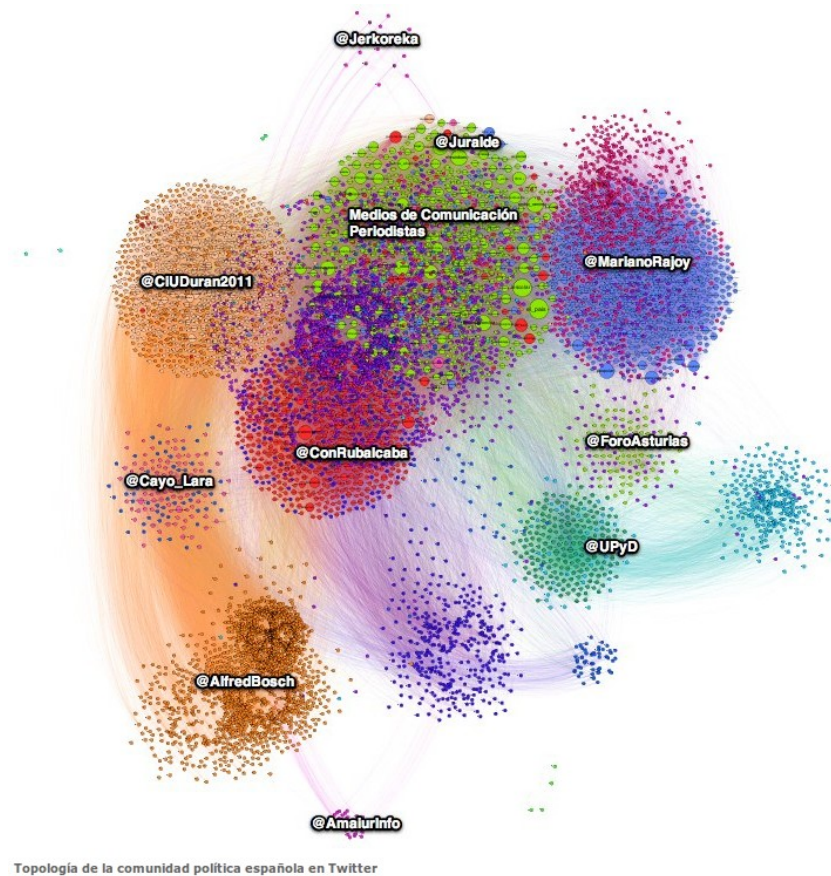


Fig 4. Topología de la comunidad política española en Twitter

Un primer resultado que llama la atención es que los periodistas y medios de comunicación ocupan el centro gravitatorio del espectro político español en Twitter. Es la de color verde y la que ocupa una posición de centralidad respecto del resto.

Los amigos de [@MarianoRajoy](#), [@ConRubalcaba](#) y [@CiUDuran2011](#) son las tres comunidades con más cercanía a los medios de comunicación. Las tres son comunidades cohesionadas y densamente relacionadas por lo que la respuesta generada por ellas seguramente será alta.

En la parte superior se encuentran los amigos de [@Jerkoreka](#). Los lazos de unión con medios y con amigos de [@CiUDuran2011](#) son los que determinan su posición. De la misma manera, que en la parte inferior del mapa se podemos encontrar a los amigos de [@AmaiurInfo](#), ya que es la zona con más proximidad a la comunidad de amigos de [@AlfredBosch](#).

La comunidad [@ForoAsturias](#), bastante dispersa, gravita en torno a los amigos de [@MarianoRajoy](#). Y entre esta comunidad y la de los amigos de [@ConRubalcaba](#) podemos ver a [@UpyD](#), aunque bastante alejada del grupo de periodistas y medios de comunicación.

Las comunidades que son satélites de [@ConRubalcaba](#) (color lila, parte inferior del mapa) y la de [@UpyD](#) (color verde-azulado) son usuarios que no están tan fuertemente interrelacionados con el grupo principal de sus comunidades, aunque revisando algunos de sus usuarios se pueden encontrar incluso canales oficiales locales. Hecho que nos indica su cercanía ideológica.

Un ejemplo más de comunidad satélite es la que se puede observar en la parte superior de la de [@MarianoRajoy](#) de color magenta. Algunos son usuarios pertenecientes a nngg o canales oficiales del [@Ppopular](#) locales. Su menor relación con el grupo principal hace que tengan esa posición satélite. Aunque en este caso menos pronunciada que la de [@ConRubalcaba](#) y [@UpyD](#).

Los amigos de [@Juralde](#) están integrados en el grupo principal de los medios de comunicación. Se puede decir que son los que están más interrelacionados con ellos.

Y los amigos de [@Cayo_Lara](#) tienen esa posición en el extra-radio debido a que usuarios como [@AcampadaSol](#) y [@DemocraciaReal](#) son una buena parte de su comunidad.

Con lo anterior, los 20 usuarios más relevantes son:

Por centralidad	Por intermediación (betweeness) Nodos clave en el flujo de información
@anapastor_tve @el_pais @iescolar @elmundoes @desdelamoncloa @pedroj_ramirez @jesusmarana @julia_otero @vicentevallestv	@iescolar @xaviertrias @danielurena @albert_rivera @publico_es @ppopular @enric_hernandez @illorente @psoe

@ppopular	@radiocable
@juanralucas	@melchormiralles
@elenaalenciano	@pedro_castro
@carlosecue	@mareamagenta
@fgarea	@garcia_abadillo
@esperanzaguirre	@julia_otero
@rtve	@jordibaste
@chdemiguel	@francesco
@arsenioescolar	@agusticolomines
@la_ser	@cyberelector
@ahernandovera	@germanyankke

Tabla 3. Usuarios más relevantes por centralidad/intermediación

6 Conclusiones y trabajo futuro

Twitter permite construir una red social en base a los conceptos de seguidores y seguidos, es decir, a diferencia de Facebook, es un grafo dirigido donde cada uno elige a quién sigue pero no elige quién le sigue.

El trabajo de investigación ha consistido en primer lugar en dar un marco teórico y de experimentación que avale la idoneidad del planteamiento, y en segundo lugar en obtener unos resultados experimentales a partir de un ejemplo real que ha permitido la contrastación empírica del estudio.

Hemos aplicado metodologías bien conocidas dentro del campo del análisis de red social para determinar la topología de la comunidad política española en Twitter como ejemplo particular en el campo de la recuperación de información y el modelado de influenciadores.

Los resultados muestran aspectos interesantes como la gran influencia de los medios de comunicación, y permiten sospechar que la metodología utilizada puede ser extrapolable y aplicable a otras situaciones diferentes.

Pero Twitter también permite construir una red social en base a lo que se habla y quién retuitea a quién. Por ejemplo, podemos seguir a personas que sean muy famosas bien por esa fama o bien porque nos haga gracia lo que dicen (pe. Lady Gaga ó Santiago Segura), pero que realmente no tienen influencia directa en los temas que nos interesan (pe. política) porque lejos de ese interés ocioso no compartimos más intereses con ellos, por lo que raramente lejos de la pura curiosidad mantendremos o generaremos conversación entendida ésta como la mención o el “retuiteo” (reenviar algo que nos ha llegado) de algo que hayan dicho dichos personajes.

Nace así otro concepto que es el de la red social que se forma de manera dinámica en torno a un determinado tema a partir de que es generado y en función de la gente que lo retuitea y a quién le retuitean. Así por ejemplo, si alguien habla de una manifestación en contra de la actual democracia española (15m, Democraciarealya, noles-votes... ver gráfico 1) y la gente empieza a retuitearlo, se están generando dos cosas, por un lado un mensaje vírico, y por otro una red social de influencia en torno a ese mensaje.

Tomando en cuenta lo anterior sería de interés incorporar como trabajo futuro el análisis de red a partir de la red creada por los RT realizados en torno a un tema, enlazando así el ARS con el PLN, y permitiendo el estudio comparativo de los indicadores clave del ARS entre la red posible y la red real, lo que permitiría determinar los temas para los que un usuario influyente es realmente influyente, o lo que es la aplicación inversa, para un determinado tema, quien es influyente y en qué sentido.

Agradecimientos

El trabajo de investigación para la consecución de Cosmos ha sido parcialmente financiado por los proyectos del ministerio ITC/464/2008 y TSI-020100-2011-56

Referencias

1. Brin, S.; Page, L. The Anatomy of a Large-Scale Hypertextual Web Search Engine. Computer Science Department, Stanford University, 1998
2. Dumazedier, J. La televisión en sociedad. Quito, Ecuador. Ciespal, 74 p., PN1992.6 D85. Centro Gumilla, Biblioteca UCAB, 1967
3. Emery, W. E. Las comunicaciones en el mundo actual. Colombia. Editorial Norma, 2003
4. Euler, L. Solutio problematis ad geometrian situs pertinentis. *Comment. Acad. Sci. U. Petrop* 8, 128-40, 1736 (Reimpreso en Opera Omnia Series Prima, Vol. 7. pp. 1-10, 1766)
5. Kleinberg, J. Authoritative sources in a hyperlinked environment. New York. In *Proc. Ninth Ann. ACM-SIAM Symp. Discrete Algorithms*, pages 668-677, ACM Press, 1998
6. Kleinberg, J. Authoritative sources in a hyperlinked environment. *Journal of the Association for Computing Machinery*, 1999
7. Kleinberg, J. Navigation in a small world. Nature Magazine, 2000
8. Molina, J. L. La ciencia de las redes. España. Asociación para el Avance de la Ciencia y la Tecnología en España (AACTE), 2004
9. Scott, J. Social Network Analysis. A Handbook. Sage Publications, 1991
10. Steinberg, Charles. Los medios de la comunicación social. México. Original Wraps, 1969
11. Verd Pericás, J.M.; Martí Olivé, J. Muestreo y Recogida de Datos en el Análisis de Redes Sociales. España. *Qüestió*, vol 23, 3. p. 507-524. 199, 1999
12. Laumann, E.O.; Marsden, P.V.; Prensky, D. The Boundary Specification Problem in Network Analysis. Applied Network Analysis, Beverly Hills. Sage. 1983
13. Blondel, V.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast Unfolding of Communities in Large Networks. Physics and Society. J. Stat. Mech. P.10008. Cornell University Library, 2008
14. Brand, Edison; Gómez Henry. Análisis de Redes Sociales como Metodología de Investigación. Elementos Básicos y Aplicación
15. Hanneman, R. A. Introducción a los Métodos del Análisis de Redes Sociales. Capítulo I.
16. Sanz Menéndez, Luis. Análisis de Redes Sociales: o cómo representar las estructuras sociales subyacentes. Apuntes de Ciencia y Tecnología Nº 7. Junio 2003