

Desvelando climas de opinión por medio del Social Media Mining y Análisis de Redes Sociales en Twitter. El caso de los Common Core State Standards.

Miguel del Fresno García¹

National Distance Education University (UNED), 28040, Madrid, Spain

Alan J. Daly

University of California San Diego (UCSD), San Diego, CA 92093, USA

Jonathan Supovitz

University of Pennsylvania (UPENN), Philadelphia, PA 19104, USA

Resumen

Los cambios derivados de la innovación tecnológica gracias a Internet y sus medios sociales están modificando la estructura de las relaciones interpersonales. El Social Media Mining (SMM) es el proceso de extraer, almacenar, representar, visualizar y analizar datos masivos generados por los usuarios con el objetivo de descubrir patrones significativos a partir de las interacciones sociales en los medios sociales de Internet. La intersección del Análisis de Redes Sociales (ARS) y SMM, en este caso Twitter por medio de sus conexiones explícitas significativas, permite hacer visible lo que hasta ahora permanecía invisible: la estructura en red de climas de opinión y la identificación de aquellos nodos o *superhubs* que tienen una influencia social desproporcionada en su red. Se presenta el análisis del clima de opinión en Twitter alrededor de los Common Core State Standards (CCSS), uno de las reformas educativas más ambiciosos de los últimos 50 años en EEUU.

Palabras clave: Minería social media-Medios sociales-Twitter-Análisis de redes sociales-ARS-Common Core State Standards.

Abstract

The changes resulting from technological innovation through the Internet and social media are modifying the structure of interpersonal relationships. Social media mining (SMM) is the process of extracting, storing, representing, visualising, and analyzing massive amounts of user generated data in order to uncover meaningful patterns from interactions within social media. The intersection of Social Network Analysis (SNA) and SMM, in this case with Twitter, allows the visualization of what was previously invisible; the network structure of climates of opinion and the key actors or *superhubs* within that structure that hold disproportionate social influence. We present an analysis of the climate of opinion on Twitter about the Common Core State Standards (CCSS), one of the most ambitious educational reforms over the last 50 years in the US.

Key words: Social media mining-social media-Twitter-Social network analysis-SNA- Common Core State Standards.

¹Enviar corespondencia a: Miguel del Fresno. UNED. Obispo Trejo s/n. Madrid 28040, Spain email: mdelfresno@der.uned.es



Introducción

La emergencia de una estructura social en red en la última década del siglo XX acelerada por la evolución de las tecnologías de la información y, en especial, de Internet ha dado lugar a la plena eclosión de lo que se ha denominado como *Era de la Información* (Castells, 2001) o *Sociedad de la Información* (Mattelart, 2002).

Internet y su arquitectura, aún heredera de su diseño conceptual al inicio de los años 90 del siglo pasado, han permitido su evolución como web 2.0 y el desarrollo y uso masivo de plataformas sociales, conocidas como *medios sociales* o *social media*. Los medios sociales son plataformas tecnológicas online creadas para facilitar diferentes formas de interacción personal -síncrona y asíncrona- con un alcance global y/o local. Lo que supone una compleja interrelación entre prácticas comunicativas, comportamiento social y tecnología, sin precedentes en la historia de la Humanidad.

Los medios sociales permiten que las personas se conecten, compartan y discutan cualquier tema, en cualquier lugar y en cualquier momento, lo que ha sido denominado como "autocomunicación de masas" (Castells, 2009:29). Fenómeno que a su vez ha experimentado una aceleración agregada, desde finales de la primera década de este siglo, debido a la penetración social de los teléfonos móviles inteligentes o *smart phones*. Así, a lo que asistimos es a una nueva expresión de cómo los cambios derivados de la innovación tecnológica modifican la estructura de las relaciones interpersonales y a una creciente *tecnologización de la realidad* (Virilio, 1991).

La concurrencia de sociabilidad y tecnología lleva a que vivamos inmersos en un denso entramado social de interacciones que forman de manera agregada una vasta red social que "nos conecta a personas, información, eventos y lugares facilitando o restringiendo el flujo de información, ideas y percepciones, en un instantáneo y masivo sistema de comunicación en red" (del Fresno, 2014:246). El hecho de que hoy todo tienda a estar conectado provoca que "ya no es ni significativa ni oportuna una separación ontológica ni fenomenológica entre el mundo online y offline" (del Fresno, 2011b:61), o *virtual* versus *real*, puesto que las prácticas sociales y comunicativas de las personas se han hibridado en un único contexto "como parte un mismo *continuum social*" (del Fresno, 2014:247). No obstante, las relaciones interpersonales que se reproducen de forma permanente y masiva en los medios sociales tienden a permanecer invisibles.

El objetivo de esta investigación es mostrar el método de cómo hacer visibles esas relaciones masivas en red sobre *cualquier tema, en cualquier lugar y en cualquier momento* por medio de *Social Media Mining* (SMM) y el Análisis de Redes Sociales (ARS). Así, es posible capturar, representar, visualizar y analizar todas esas interacciones en red en una escala que no era posible antes y hacerlo dentro de uno de los medios sociales con mayor crecimiento e importancia: Twitter. Para ello se eligió como objeto de investigación el debate alrededor de los *Common Core State Standards* (CCSS), uno de los proyectos educativos más ambiciosos de los últimos 50 años en EEUU, identificando a los nodos (individuos, grupos, instituciones, etc.) con mayor influencia en cada red analizada.

Social media, Social media mining y Análisis de redes sociales

Los medios sociales son una muestra más de la extraordinaria habilidad de las personas para generar, difundir e intercambiar significados en la comunicación interpersonal colectiva, en un sistema masivo en red y en tiempo real donde todo tiende a estar conectado.

Los medios sociales se pueden organizar en diferentes categorías donde cabe mencionar, sin ser las únicas, plataformas de *social networking* (Facebook, LinkedIn), *microblogging* (Twitter, Weibo), fotografía (Flickr, Instagram, Pinterest), vídeo (Youtube, Vimeo, MetaCafe), *social news* (Meneame, Digg, Reddit), emisión en directo (Livecast, Ustream), *social gaming* (World of Craft), *bookmarking* (Delicious, StumbleUpon), blogs (Wordpress, Blogger), etc. Los medios sociales se han convertido en fuente de una ingente cantidad de datos que producen de forma ubicua, incesante y masiva, lo que se denomina *Big Data* (Mayer-Schönberger, Cukier 2014). Este tipo de datos ofrece oportunidades y desafíos a los investigadores, donde lo más interesante reside, no en las cantidades de datos, sino en qué podemos hacer con estas grandes cantidades de datos que no se puede hacer con pequeñas cantidades.

Recabar y proteger datos ha sido una constante para poder recuperarlos, analizarlos y generar así información y conocimiento. Con los medios sociales nos enfrentamos a la misma necesidad con una diferencia significativa: la *velocidad* a la que se producen nuevos datos², lo que anticipó el sociólogo Paul Virilio al afirmar

² IBM estima que "cada día, se generan 2,5 trillones de bytes de datos -tanto que el 90% de los datos en el mundo hoy se han generado en los últimos dos años. Estos datos provienen de todas partes: sensores utilizados para recopilar información sobre el clima, mensajes en medios sociales, imágenes y vídeos, registros de transacciones de compra y señales de teléfono móvil, GPS por nombrar algunos". Datashift,

que “el tiempo real prevalece sobre el espacio real y la geosfera. La supremacía del tiempo real, la inmediatez, sobre espacio y superficie es un hecho consumado y tiene un valor inaugural (anuncia una nueva época)” (Virilio, 1995) por lo que “la emergencia de medios sociales hiper-inmediatos resalta la necesidad de nuevas formas de investigación en tiempo real” (Elmer, 2013:19).

El *Social Media Mining* (SMM) (Zafarani et al., 2014; Guandong et al. 2013; Jinhui et al., 2012) se puede definir como el proceso de extraer, almacenar, representar, visualizar y analizar los datos generados por los usuarios con el objetivo de descubrir patrones significativos (difusión de información o rumores, influencia, homofilia, comportamiento social o de consumo, predicción, etc.) a partir de las interacciones sociales en los medios sociales de Internet. Profundizar en la comprensión, real y potencial, tanto de los patrones de interacciones como de la identificación de aquellos nodos con una influencia desproporcionada o *superhubs* dentro de la red de estudio y sus efectos, lo que se ha convertido en un tema cada vez más significativo para más campos del conocimiento.

El SMM suma una mayor complejidad a la captura de datos tradicional ya que los datos de los medios sociales son: a) poco homogéneos, b) poco estructurados (las formas de los contenidos dependen de cada medio), c) masivos y se producen sin fin en cualquier momento y lugar y d) están cargados de formas muy diversas de relaciones sociales (amigos, seguidores, seguidos, lectores, etc.) más o menos significativas para la investigación.

De igual forma los medios sociales presentan características novedosas que incrementan el problema de la captura, representación, análisis y su comprensión como fuentes:

- 1) los datos de los medios sociales son generados por la actividad de las personas en su despliegue social online por medio de las múltiples tipos de relaciones que se establecen (compartir imágenes, ideas, textos, vídeos, etc.), por lo que estamos ante datos multidimensionales.
- 2) los datos crecen, se multiplican y desaparecen cada día, por lo que las fuentes no son estables y supone una exigencia constante, técnica y metodológica, el mantener identificadas y activas las tecnologías de captura de esas fuentes para no perder el flujo de datos.

dedicada a comercializar datos sociales de gran cantidad de fuentes sostiene p.e. que se generan 500 millones de tweets diarios, más de 50 millones de interacciones cada día en Facebook o más de 100 millones de post diarios en Sina Weibo <http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html> Consulta [30-09-14]

3) la heterogeneidad de las fuentes dificulta la agregación de los datos de forma consistente en la búsqueda e identificación de patrones significativos.

La tarea del SMM no se reduce a una cuestión de qué tecnología de rastreo elegir (siempre limitadas por sus creadores y usuarios) sino de un proceso más general de integración de tecnología, análisis y comprensión en una emergente y nueva área de conocimiento. Así, la intersección del ARS y el SMM permite capturar, representar y analizar esas interacciones masivas, sean consecuencia de la acción individual o colectiva, como redes con una profundidad y escala nunca antes posible. Lo que a su vez presenta dos hechos originales: 1) hacer visible lo que hasta ahora era invisible: la estructura de las relaciones en redes en "un lugar sin espacio" (del Fresno, 2011b:19), el ciberespacio y 2) crear un artefacto cognitivo para amplificar nuestro conocimiento sobre la complejidad de la actividad humana cotidiana en su despliegue local y global gracias a Internet³. Una de las plataformas sociales más destacadas por su dinamismo y crecimiento, local y global, es Twitter⁴.

Twitter como medio de intersección de todos los medios

Twitter (www.twitter.com) es un medio social online global y gratuito donde los usuarios escriben mensajes cortos limitados a 140 caracteres, conocidos como *tweets*, que se envían a todos aquellos (*followers*) que han optado por recibir los tweets del emisor. Al publicar un tweet (vía web, sms, ordenador, tablet o teléfono móvil) es posible incrustar además de texto, imágenes, enlaces a vídeo o cualquier otro medio social o profesional con presencia online y *hashtags* (una palabra o frase prefijada con el símbolo # que las convierte en metadatos).

Twitter es en la actualidad el medio social más rápido, simple y económico por donde circula todo tipo de informaciones, noticias, ideas, eventos, rumores, materiales multimedia, etc. emitido desde cualquier medio profesional u otros medios sociales en tiempo real. Así, en Twitter interseccionan tanto medios profesionales (televisión, prensa escrita, revistas, radio, etc.) como cualquier otro medio social (Facebook, Instagram, Youtube, Flickr, blogs, foros, etc.) de un muy amplio registro de tipologías de perfiles de usuarios (individuos, ONGs, entidades gubernamentales, medios de comunicación de masas, grupos de presión, revistas

³ Internet World Stats: Usage and Population Statistics <http://www.internetworldstats.com> Consulta [30-09-14]

⁴ Twitter Reports Second Quarter 2014 Results <https://investor.twitterinc.com/releasedetail.cfm?ReleaseID=862505> Consulta [30-09-14]

científicas, empresas, marcas, etc.). Todo lo cual otorga a Twitter un campo privilegiado para la investigación ya que, en la práctica, estaría desempeñando el rol de una suerte de sistema nervioso central de Internet, esto es, un medio de intersección de “*every media & medium*” (Dorsey, 2012).

La intersección del SMM, ARS y Twitter, como se presenta en esta investigación, hace posible rastrear, capturar y analizar enormes o pequeñas cantidades de datos –dependiendo de la pregunta de investigación– tanto en tiempo real como de forma retrospectiva y su representación y análisis como redes, lo que supone un acceso sin precedentes a los registros de la actividad humana online en el tiempo⁵ y la visibilización de las relaciones en red en Twitter que hasta ahora eran invisibles.

Relaciones explícitas significativas

En los medios sociales, la interpretación y la mediación de sentido se producen en el nivel de interacción, y estas interacciones representan relaciones explícitas que unen a las personas. Así los usuarios de manera intencional y consciente crean *conexiones explícitas* en los medios sociales, mientras que las *conexiones implícitas* se infieren de sus comportamientos online (Golbeck, 2013:16).

Más allá de cómo fue ideado Twitter en su origen, en 2006, los usuarios han desarrollado “formas creativas para conseguir el máximo rendimiento de cada tweet mediante el uso de diferentes herramientas de comunicación” (Lovejoy et al. 2012:313). La forma más común de relación explícita en los medios sociales es la *amistad* o *friending* recíproco (p.e. en Facebook), aunque no siempre resulta significativa para el análisis y la investigación porque está basada en una conexión hipotética que puede estar o no activa en la práctica y porque no todos los medios sociales imponen la reciprocidad del *friending* como obligatoria (p.e. Twitter). Otras formas de relaciones explícitas serían *likes*, *follows*, *blocks*, *favorites*, *listing*, *emails*, *mensajes instantáneos*, *pokes*, *+1*, *retweets*, *mentions*, *answers*, etc. dependiendo de cada medio social. Las relaciones explícitas o su ausencia (p.e. *dislikes* en Facebook) viene definida por la experiencia de usuario creada por el código con el que está programada cada plataforma. En Internet “el código es la ley del ciberespacio” (Mitchell, 1995:111), la forma en que se facilitan o restringen las prácticas comunicativas, en definitiva, las relaciones explícitas. En Twitter, Facebook o, más en general, en Internet “el lugar de estudio no es el objeto de

⁵ Una empresa privada, Datasift comercializa los datos históricos de la actividad en los medios sociales (<http://datasift.com/platform/historics>) Consulta [30-09-14]

estudio" (del Fresno, 2011b:21) sino las prácticas sociales derivadas de la comunicación interpersonal colectiva y sus efectos en los comportamientos; y para abordar este reto es necesario el ARS y su corpus de conocimiento.

Desde el punto de vista del ARS el aspecto más significativo de Twitter para la investigación es que facilita: 1) la comunicación e intercambio de información masiva e interpersonal de manera simultánea y 2) al menos dos tipos de relaciones explícitas significativas y diferenciales: *retweets* (RTs) y *menciones o respuestas* (MTs), claves en el método de representación en red propuesto. Estos dos tipos de relaciones explícitas son p.e. más significativas que las relaciones de seguidores-seguidos (*followers-following*) ya que RTs y MTs son conjuntos de conexiones explícitas superpuestas en la misma red.



Figura 1: RT o retweet y uso de hashtag



Figura 2: MT o mención-respuesta y uso de hashtag

MTs y, en especial, RTs son los dos actos comunicativos explícitos más significativos sobre el resto de posibles elecciones en Twitter puesto que "investigaciones preliminares han encontrado que sólo el 6% de todos los tweets son retweeteados" (Elmer, 2013: 19) y "menos de 1 de cada 200 mensajes es retweeteado después de su primera hora de vida" (Geere, 2010). Así la clave del método propuesto reside en entender las conexiones explícitas desde el punto de vista del ARS de la siguiente forma:

1. Los RTs son el número de veces que el tweet de un individuo "RT @nombreusuario" (figura 1) es reenviado al *timeline* de todos los seguidores del usuario que hace el RT, por lo que el recibir un alto número de RTs, entendido como in-degree, puede inferirse como la habilidad de un sujeto en su rol de productor-emisor (micromedio) de información significativa para terceros y puede ser utilizado como una medida de influencia.

2. Las MTs (menciones y respuestas) son el número de veces que un individuo aparece en tweets de terceros como "@nombreusuario" (figura 2), por lo que el número de MTs, entendido como in-degree, puede inferirse como la habilidad de un sujeto para establecer relaciones dialógicas con otros y puede ser utilizado como una medida de influencia⁶.
3. Finalmente, cabe la existencia de individuos (nodos) que destaquen en ambos tipos de relaciones explícitas, esto es que reúnan las habilidades de grandes productores-emisores de información circulando por la red y, al mismo tiempo, con altas habilidades relacionales capaces de establecer relaciones dialógicas con un gran número de nodos en su red.

Un alto número de RTs mostraría individuos con mayores habilidades para alcanzar cuotas significativas de difusión de sus mensajes, con una alta diversidad estructural en su red y que tienden a maximizar su influencia en la comunicación evitando la formación de *structural holes*.

Un alto número de MTs mostraría individuos con más habilidades relacionales o de *engagement*, con una alta diversidad estructural en su red que consiguen tender a maximizar la eficacia en su relación dialógica con terceros.

El objetivo de la investigación es la propuesta de identificación de *superhubs* en grandes redes extraídas de Twitter a partir de sus relaciones explícitas más destacadas⁷.

El objeto de estudio: Common Core State Standards

Al presentar un nuevo método que une ARS, SMM y Twitter era necesario elegir un objeto de estudio que cumpliera dos condiciones: 1) ser lo más significativo posible en su repercusión social sin tener de forma necesaria escala global y 2) que el clima de opinión estuviera activo a lo largo del tiempo.

El trabajo de campo consistió en rastrear, capturar, almacenar, representar y analizar la información generada en Twitter alrededor de los *Common Core State*

⁶ Para evitar que algunos perfiles que no participan en el clima de opinión pero que son mencionados en numerosas ocasiones aparecieran con un alto in-degree, lo que a lo largo del estudio denominamos como *Efecto @BarackObama* se decidió incluir sólo como nodos activos de la red a aquellos que al menos hubiesen emitido un tweet con las keywords que definen el clima de opinión.

⁷ Para estos objetivos se ha desarrollado la herramienta online Social Runner Lab (www.socialrunnerlab.com) una tecnología online específica para la captura, visualización y representación en tiempo real de climas de opinión de cualquier volumen en Twitter a partir de búsquedas complejas que la realizada con CCSS.

*Standards (CCSS)*⁸ o *Normas Estatales Comunes*, uno de los proyectos educativos más ambiciosos de los últimos 50 años en EEUU, para identificar a los nodos ejerciendo una influencia desproporcionada en el clima de opinión alrededor de los CCSS.

El contexto social de los CCSS

En EEUU es inusual la existencia de normas nacionales ya que cada estado tradicionalmente establecía la definición de sus propios estándares y habilidades exigibles a los estudiantes. Esto está cambiando debido a la aplicación de los CCSS. Los CCSS en matemáticas, artes y lengua inglesa (ELA) fueron desarrollados por la *National Governors Association* y el *Council of Chief State School Officers* y han sido adoptadas hasta la actualidad por 46 Estados. Los CCSS tienen como objetivo definir los conocimientos que los estudiantes deben tener en los niveles de Primaria y Secundaria antes de llegar a la Universidad⁹. Los CCSS requerirán de los diferentes estados y distritos la reorientación, y en muchos casos, una revisión profunda del currículo, enseñanza y evaluaciones. Como tal, los CCSS suponen un cambio de política en la educación de gran calado en términos de su amplitud y profundidad.

A pesar de que en su origen, 2012 e inicios de 2013, el debate alrededor de los CCS era eminentemente técnico entre especialistas relacionados con la Educación, desde el verano de 2013 y durante 2014 el debate se fue transformando y adquiriendo una enorme repercusión social, cultural y política. La oportunidad para estudiar desde el punto de vista del ARS los CCSS fue triple: 1) un debate de gran alcance limitado a un idioma y un país 2) confirmar, tras unos primeros análisis, que Twitter estaba siendo utilizado como un medio de intersección privilegiado del debate político, social y cultural alrededor de los CCSS y 3) que durante el primer semestre de 2013 identificamos un déficit de trabajos científicos alrededor de los CCSS.

⁸ Common Core se presenta como: "State education chiefs and governors in 48 states came together to develop the Common Core, a set of clear college- and career-ready standards for kindergarten through 12th grade in English language arts/literacy and mathematics. Today, 43 states have voluntarily adopted and are working to implement the standards, which are designed to ensure that students graduating from high school are prepared to take credit bearing introductory courses in two- or four-year college programs or enter the workforce". En <http://www.corestandards.org/about-the-standards/frequently-asked-questions>) Consulta [30-09-14]

⁹ Ver Common Core State Standards Initiative (2013). <http://www.corestandards.org> Consulta [30-09-14]

Así, se fijaron los siguientes objetivos de investigación a partir de las conexiones explícitas definidas:

- 1) representar la red completa e identificar que porcentaje de esa red representaba el Componente Gigante (CG).
- 2) representar las comunidades estructurales existentes dentro del CG.
- 3) representar la red del 1% del total de la red tanto por in-degree como por out-degree e identificar a los *superhubs* dentro de esas redes por ambas métricas.
- 4) representar la red del 0,25% del total de la red tanto por in-degree como por out-degree e identificar a los *superhubs* dentro de esas redes por ambas métricas.
- 5) representar de forma geolocalizada la actividad de aquellos usuarios que aportan su localización en modo público tanto por in-degree como por out-degree.

Método

La captura y extracción de los datos presentados en esta investigación se realizó directamente desde la API (*application programming interface*) de Twitter durante seis meses, desde el 1 de septiembre de 2013 al 28 de febrero de 2014¹⁰. Tras la definición del objeto de estudio por medio de la palabra clave con o sin *hashtag*, "CommonCore" o "Common Core", se capturaron 183.346 tweets de los cuales 89.450 (48,7%) fueron identificados como RTs y 129.339 MTs (gráfico 1).

¹⁰ El proceso de captura ha seguido activo hasta finales de 2014 para recoger el efecto de las elecciones del 4 de noviembre de 2014 donde total de 471 escaños fueron elegidos: 36 escaños en el Senado, incluyendo tres elecciones especiales, y los 435 escaños del Congreso. El clima de opinión de los CCSS a partir del cambio político significativo, por el que los republicanos consiguieron mayoría en el Congreso se realiza con SocialRunnerLab (<https://socialrunnerlab.com>) una tecnología creada por Miguel del Fresno y Alan J. Daly para la captura y representación de climas de opinión múltiples en tiempo real en Twitter.

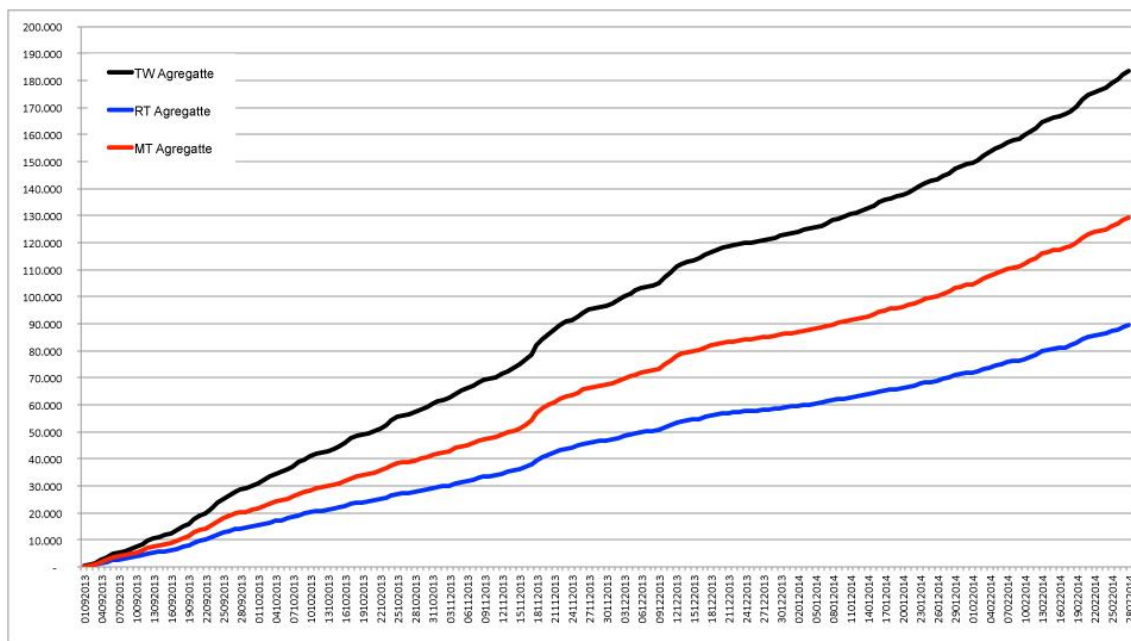


Gráfico 1. Evolución acumulada de tweets, MTs y RTs

A partir de estos datos se generaron los siguientes ficheros:

- 1) red completa (una red agregada con RTs y MTs, las redes presentadas en este artículo),
- 2) red de RTs y red de MTs por separado (a presentar en próximo artículo)

El formato .gexf fue el elegido para generar la red completa. Luego se importó y explotó el fichero con dos tipos de software de ARS (en Gephi¹¹ primero y después exportando a ficheros .dl para su análisis con UCINET¹² de las redes del 1% y 0.25%) para su representación, visualización y explotación analítica. Finalmente, los *nicknames* o nombres de usuarios se anonimizaron con un ID único para garantizar el anonimato.

Resultados

Con los datos de la red completa se procedió al cálculo de las métricas de *modularity*, *in-degree* y *out-degree* y sucesivos filtrados hasta llegar a la

¹¹ Gephi es un software gratuito de código abierto para la visualización interactiva, exploración y análisis de redes incluso de grandes tamaños <https://gephi.org>

¹² UCINET es un software no gratuito desarrollado por Lin Freeman, Martin Everett y Steve Borgatti para el análisis y explotación estadística de pequeñas redes que, además sólo está disponible para Windows <https://sites.google.com/site/ucinetsoftware/home>

identificación de los nodos de las redes del 1% y 0,25% por in-degree y out-degree.

- La Red Completa resultante es una red direccionada de 63.493 nodos y 125.879 relaciones que se fue filtrando para obtener las subsiguientes redes (gráfico 2).
- El Componente Gigante (CG) con 51.254 nodos (80,72% del total) y 123.035 relaciones (97,74% del total), (gráfico 3) lo que da una idea de la significativa densidad relacional dentro del clima de opinión.
- Aplicación de la métrica de modularidad al CG, (gráfico 4) que muestra la existencia, representada en colores, de tres grandes comunidades estructurales¹³.
- La red del 1% por in-degree: 643 nodos (1,01% del total); 3.811 relaciones (3,03% del total), (gráfico 5). Aplicando el procedimiento del actor clave o *key player* (Borgatti, 2006) a esta red obtenemos que se alcanzaría el 94,5% de la red con sólo 18 nodos *superhubs* (2,79%)¹⁴ lo que supone una fracción muy pequeña para conseguir el *site percolation* (Newmann, 2010:592).
- Red del 0,25% por in-degree: 163 nodos (0,25% del total); 882 relaciones (0,7 % del total), (gráfico 6). Partiendo de la hipótesis de que vacunar/infectar una pequeña fracción de nodos -en este caso *superhubs* identificados por terceros como influyentes- y sus relaciones en una red, *site percolation*, puede tener efectos muy significativos en la misma a la hora de frenar/expandir información, ideas o percepciones y provocar un efecto *knock-on* o *efecto en cadena* en la red en el que el beneficio de frenar/expandir a ese reducido número de nodos provocará cambios significativos no sólo en la estructura de toda la red sino también en el comportamiento de los individuos que la forman. Aplicando el procedimiento del *actor clave* o *key player* a esta red obtenemos que se alcanzaría el 95,6% de la red con sólo 9 nodos *superhubs* (5,5%)¹⁵ para conseguir el *site percolation*.
- Red del 1% por out-degree: 641 nodos (1,01% del total); 3.140 relaciones (2,49% del total), (gráfico 7). Aplicando el procedimiento del actor clave o *key player* a esta red obtenemos que se alcanzaría el 94,8% de la red con sólo 19

¹³ En un análisis de contenidos posterior estas comunidades estructurales han mostrado responder a tres posiciones diferentes alrededor de los CCSS, a publicar en próximos estudios

¹⁴ Con los ID 9, 95, 109, 111, 115, 145, 148, 155, 171, 193, 198, 295, 303, 321, 442, 454, 460 y 471.

¹⁵ Con los ID 44, 45, 46, 48, 55, 87, 91, 117, 126 y 145.

nodos *superhubs* (2,96%)¹⁶ lo que supone una fracción muy pequeña para conseguir el *site percolation*.

- Red del 0,25% por out-degree: 158 nodos (0,25% del total); 705 relaciones (0,56% del total), (gráfico 8). Partiendo de la misma hipótesis que en la red del 0,25% por in-degree y aplicando el procedimiento del *actor clave* o *key player* a esta red obtenemos que se alcanzaría el 95,1% de la red con sólo 5 nodos *superhubs* (3,1%)¹⁷ para conseguir el *site percolation*.
- Geolocalización por in-degree. Con aquellos individuos que hacen público en su perfil la información de su geolocalización en EEUU se procedió a la representación por su in-degree (gráfico 9), que muestra la actividad de las dos grandes comunidades estructurales identificadas a lo largo de todo EEUU y un cierto equilibrio en la distribución de la influencia de la información viralizada y del debate alrededor de los CCSS.
- Geolocalización por out-degree. Con aquellos individuos que hacen público en su perfil la información de su geolocalización en EEUU se procedió a la representación por su out-degree (gráfico 10), que muestra los principales focos de emisión, con muy pocos nodos ejerciendo en la práctica como principales *broadcasters*. Y donde se refleja el origen geográfico de la actividad de las dos grandes comunidades estructurales en las dos costas y en el centro del país del debate alrededor de los CCSS¹⁸.

¹⁶ Con los ID 9, 95, 109, 111, 115, 145, 148, 155, 171, 193, 198, 295, 303, 442, 454, 460, 476 y 543.

¹⁷ Con los ID 42, 48, 52, 83 y 119.

¹⁸ Esta investigación ha coincidido en la práctica, durante el período analizado, con el debate público con los dos principales focos de activismo a favor de los en la costa este (donde incluso partidarios de los CCSS han derivado en alguna posición crítica en New York) y oeste; y de oposición dominante en los estados centrales.

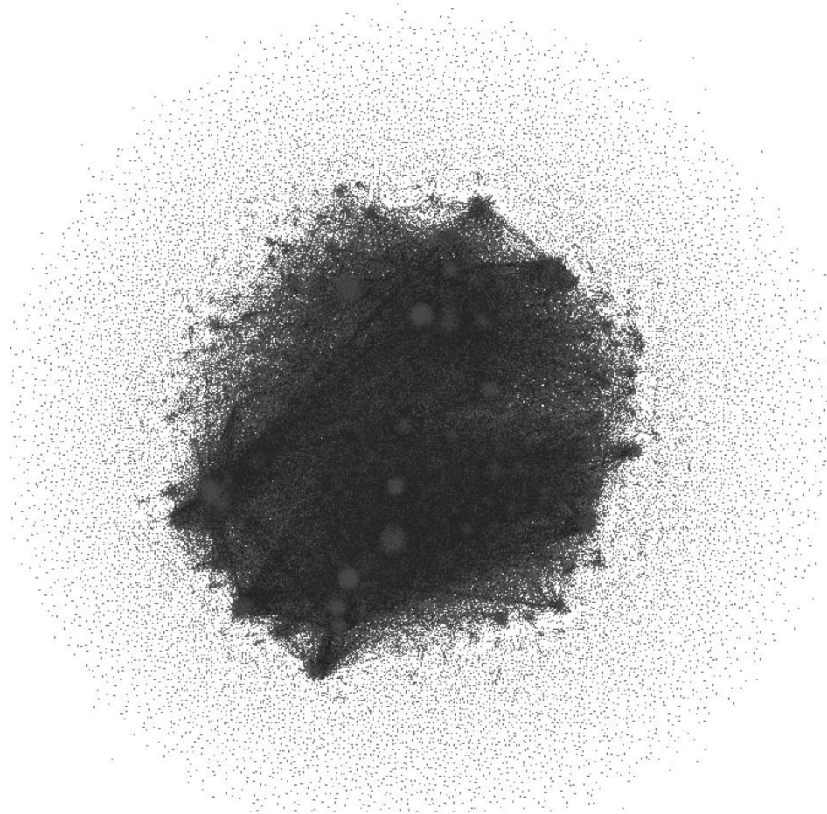


Gráfico 2. Red completa.

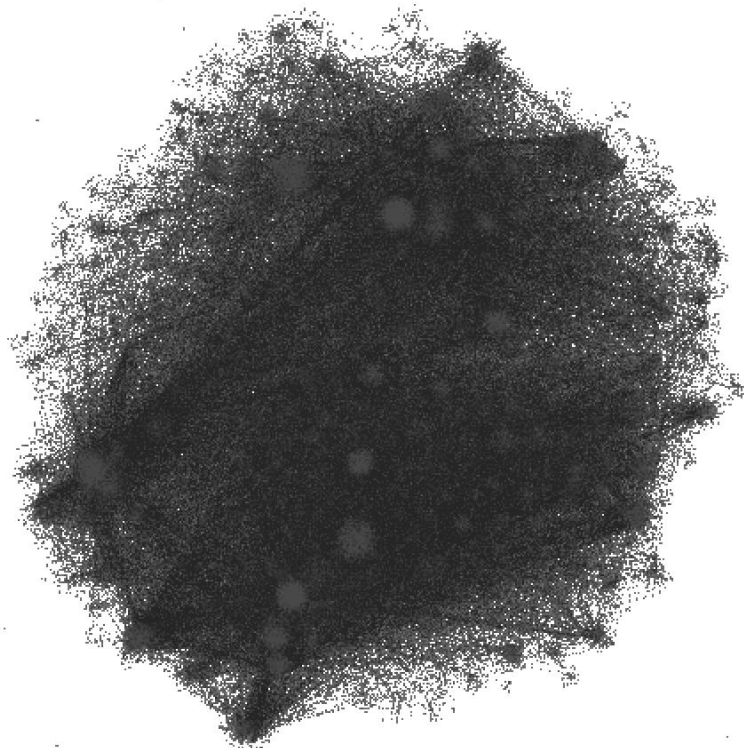


Gráfico 3. Red Componente Gigante

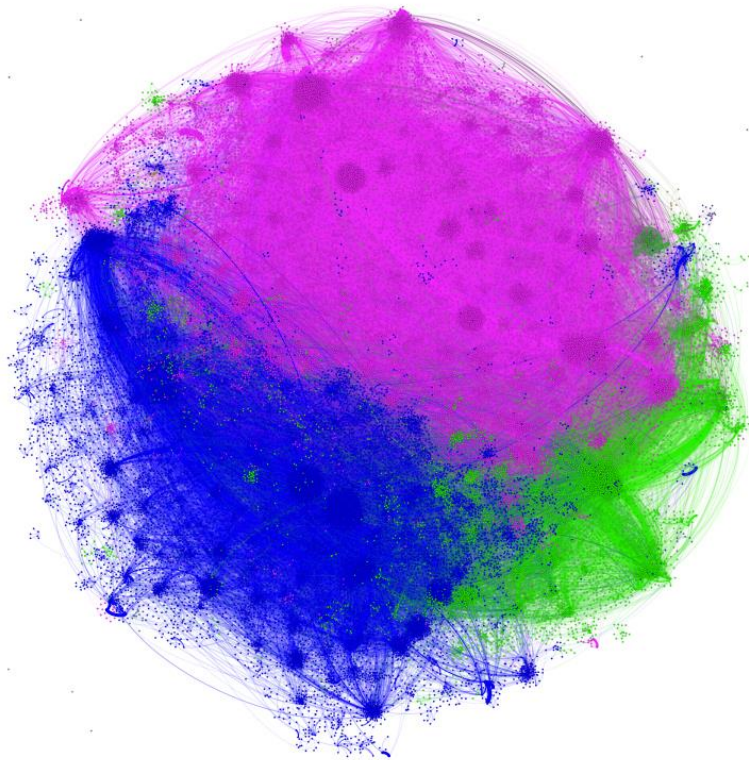


Gráfico 4. Modularidad del Componente Gigante



Gráfico 5. Red del 1% por in-degree

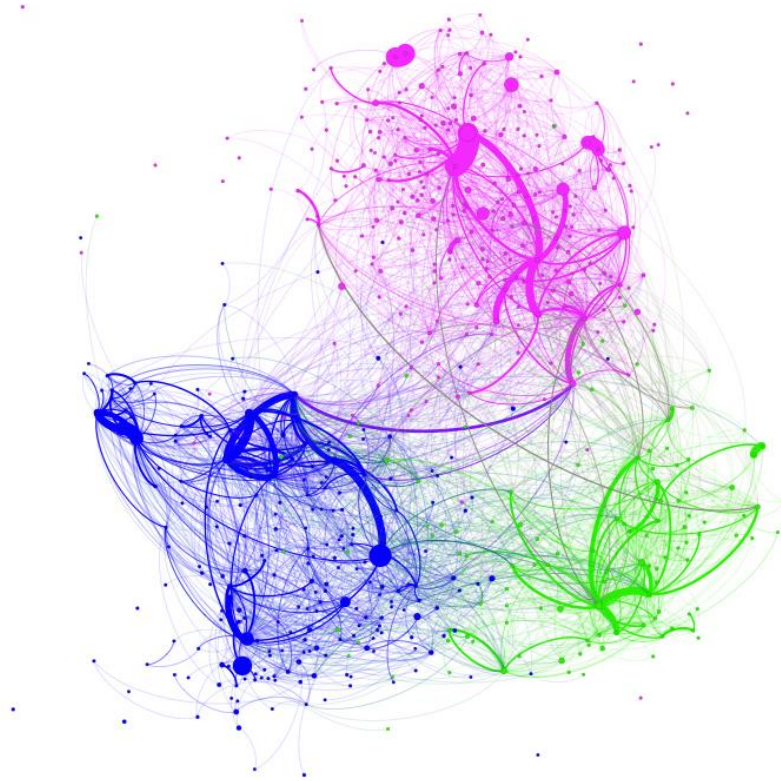


Gráfico 6. Red del 1% por out-degree



Gráfico 7. Red del 0,25% por in-degree



Gráfico 8. Red del 0,25% por out-degree

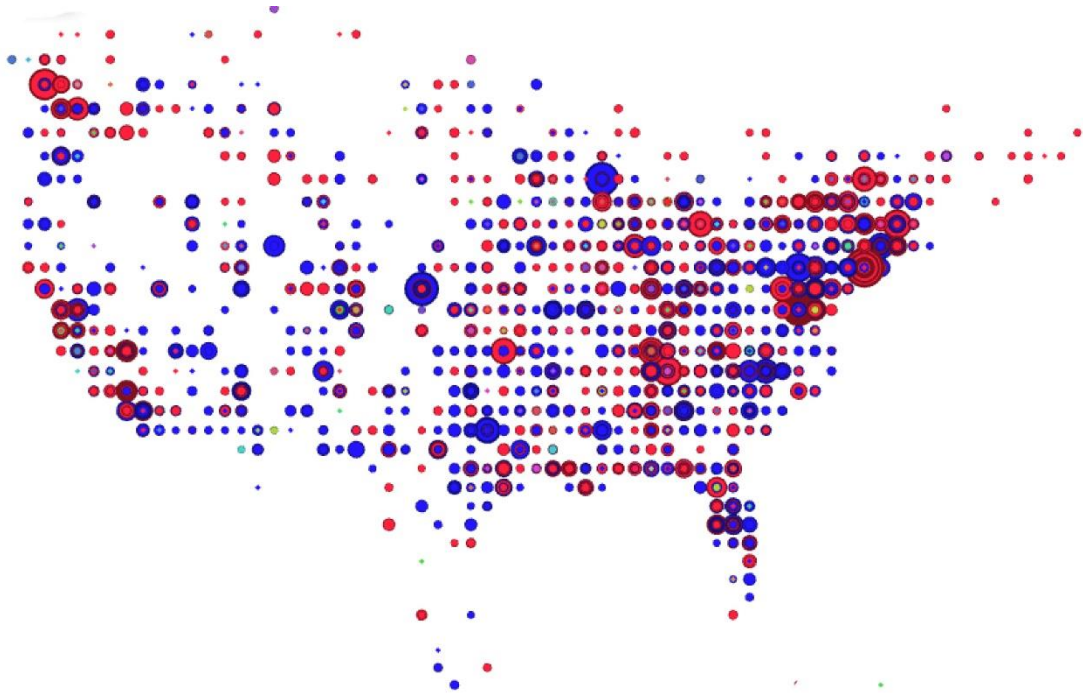


Gráfico 9. Red completa. Geolocalización por in-degree

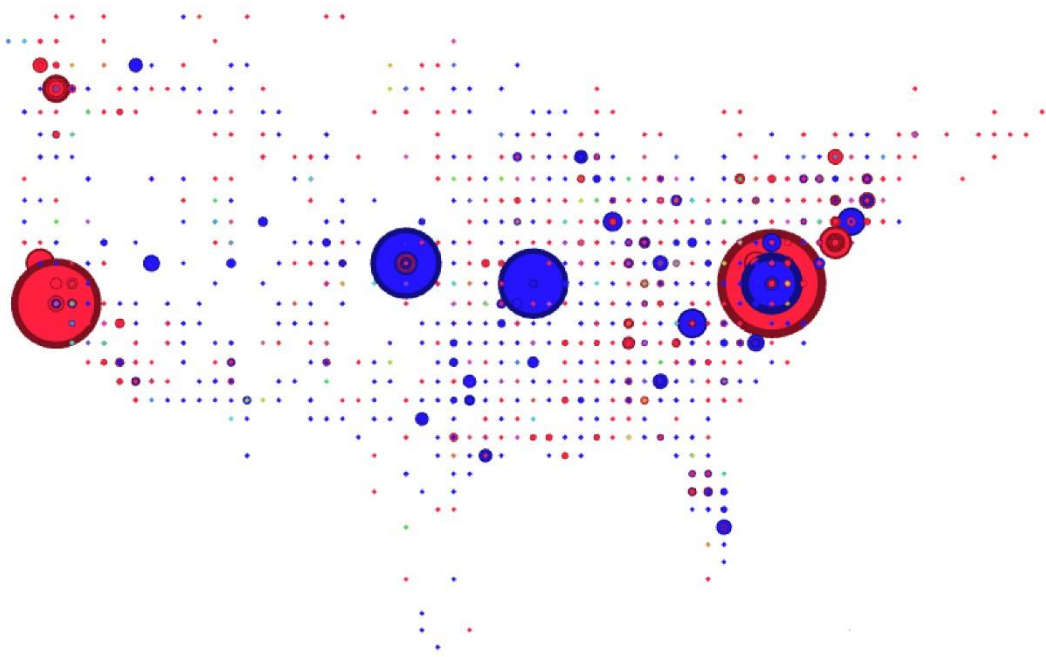


Gráfico 10. Red completa. Geolocalización por out-degree

Discusión

La intersección entre la comunicación interpersonal colectiva en Twitter, las técnicas de SMM y el ARS presenta cuatro características claves para este estudio: 1) se asume una intuición estructural de las relaciones sociales, 2) se capturan, reúnen, representan y analizan sistemáticamente datos empíricos relacionales, 3) se utilizan modelos matemáticos para el análisis junto con la tecnología y 4) se crean y comparten visualizaciones de las relaciones y los patrones de interacción, que permiten la generación de ideas estructurales significativas y su comunicación a los demás, lo que coincide de pleno con L.C. Freeman (2000, 2004) sobre el desarrollo del ARS como disciplina social.

El objetivo para estudiar redes en esta investigación es la identificación de los principales actores o *superhubs* en el debate social, cultural y político alrededor de los CCSS en Twitter, a partir de sus relaciones más significativas, utilizando como datos los generados por los propios usuarios en su comunicación. Twitter ofrece como hecho singular el estar desempeñando, en la práctica, la función de medio de intersección del resto de medios. Esto, es una suerte de columna vertebral o sistema nervioso central por donde circulan y pueden ser identificados, capturados, analizados y representados los contenidos de la comunicación interpersonal colectiva que facilita la arquitectura de Internet.

Este método aplicado a los CCSS ha ofrecido a los investigadores nuevas oportunidades para comprender las prácticas comunicativas (del Fresno, 2011a, 2012) y profundizar en la comprensión, a presentar en próximas publicaciones, de macro-reacciones colectivas a micro-acciones individuales dentro de las redes de los climas de opinión.

Existen evidencias que muestran claras conexiones entre el contagio/difusión/desarrollo de enfermedades infecciosas y la difusión de la información puesto ambas que se propagan de persona a persona a través de redes, de influencia u homofilia, que muestran gran similitud estructural (Easley & Kleinberg, 2010) lo que ha llevado a que la difusión de ideas se conceptualice como *contagio social* o *social contagion* (Burt, 1987; Sun et al. 2009) y es aplicable a los medios sociales de Internet y, de forma destacada, a Twitter.

Este tipo de datos presentan nuevas oportunidades y desafíos a los investigadores, donde lo más interesante reside, no sólo en las cantidad de datos, sino en qué podemos hacer con estas grandes cantidades de datos que no se puede hacer con pequeñas cantidades. Al menos se presentan ya dos grandes retos para los investigadores:

- 1) *El reto de la Complejidad* o cómo capturar y agregar de forma consistente los datos multidimensionales, poco homogéneos, poco estructurados, masivos, que se producen sin fin en cualquier momento o lugar y que tienen, a su vez, fuentes heterogéneas e inestables (que pueden aparecer y desaparecer) manteniendo como objetivo central la búsqueda e identificación de patrones significativos.
- 2) *El reto de N=todo* o cómo desarrollar metodologías que nos permitan trabajar con la totalidad de los datos que se producen, esto es, cómo investigar con universos completos. Si el muestreo es una técnica desarrollada para épocas de escasez de información, hace tiempo que abandonamos colectivamente esa realidad. No obstante, en época del *Big Data* o de la *Era del petabyte* que pronostica “un mundo en el que ingentes cantidades de datos y las matemáticas aplicadas reemplazan cualquier otro instrumento” lo que implica que “el volumen de los datos obviará la necesidad de teoría, e incluso del método científico” (Anderson, 2008). Mientras tanto, debemos mantener el foco en los procesos de análisis y la toma de decisiones correctas a través de la identificación de patrones significativos y siendo conscientes de que “la promesa implícita del *Big Data* es que la solución al exceso de información pasa por mayores cantidades de datos” (del Fresno, 2014:247) .

Una evidente limitación de este tipo de investigación residiría en el déficit de acceso a los significados que circulan por las redes. Para ello existen metodologías emergentes como la netnografía (del Fresno, 2011b; del Fresno & López 2014), una propuesta entre la sociología y la antropología que responde a la necesidad de aproximaciones interdisciplinarias para aunar estudio de la estructura y acceso a los significados.

En relación con los CCSS se presentará en próximas publicaciones los resultados de las dos líneas de investigación que abrió este trabajo inicial: 1) ampliar y repetir el análisis hasta los 14 meses donde se ha producido un cambio político significativo en EEUU y, aplicando el mismo método, identificar y comparar los *superhubs* en ambos cortes temporales para comprender mejor la dinámica en red alrededor de los CCSS y 2) el análisis de contenidos de los tweets alrededor de qué ideas y percepciones están difundiéndose, qué tipo de perfiles individuales, qué grupos ideológicos o qué tipos de perfiles institucionales están activos y cómo puede afectar al debate general la expansión del activismo en Twitter, a favor y en contra de los CCSS, al poder haberse convertido en una *guerra cultural* (Thomson, 2010; Zimmerman 2002) o como mínimo en una *proxy war* en EEUU. De forma paralela se han realizado entrevistas profundas a los *superhubs*, identificados en los primeros 6 meses de investigación, para conocer de primera mano sus motivaciones particulares alrededor de los CCSS.

Para los investigadores y, en especial, para aquellos centrados en la comunicación la clave tras identificar los *superhubs* es, sin duda, definir el *input* idóneo (tipo y cantidad de información) a introducir en una red con el objetivo de alcanzar el *tipping point* (Schelling, 1978; Gladwell, 2000) que permita la óptima difusión o, también, todo lo contrario. El riesgo estriba en que si una vez identificados los *superhubs* son inundados con exceso de información comenzará a declinar su eficacia (Allen, 1977) por lo que próximas investigaciones deberán afrontar este reto.

Desde el punto de vista del ARS el mayor reto sigue estando en la comprensión de las dinámicas sociales en red de los climas de opinión en Twitter, en particular, y en los medios sociales, en general, cómo evolucionan a lo largo del tiempo y cómo, en definitiva, pueden servir para comprendernos mejor como seres sociales.

Bibliografía

- Anderson, Chris (2008). "The end of theory: the data deluge makes the scientific method obsolete". *Wired*, Jun 23. Disponible en <http://mdelf.es/1dkao> Consulta [30-09-14]
- Allen, Th. J. (1977). "The role of person to person communication networks in the dissemination of industrial technology", *Sloan School of Management. Working paper 939-77 MIT*. Disponible en <http://mdelf.es/1b4xs> Consulta [30-09-14]
- Borgatti, S.P. (2006). "Identifying sets of key players in a network". *Computational, Mathematical and Organizational Theory*. 12(1), pp. 21-34 <http://dx.doi.org/10.1007/s10588-006-7084-x>
- Burt, R.S. (2000). "The network structure of social capital", en R. I. Sutton & B. M. Staw, eds. *Research in organizational behavior*. pp. 1-83. Greenwich, CT: JAI Press.
- Castells, M. (2001). *La era de la información: economía, sociedad y cultura*. Madrid: Alianza editorial.
- Castells, M. (2009). *Comunicación y poder*. Madrid: Alianza Editorial. <http://dx.doi.org/10.2307/40184158>
- Del Fresno, M. (2011a). "Infosociabilidad monitorización e investigación en la web 2.0 para la toma de decisiones", *El profesional de la información*, 20 (5), pp. 548-554 <http://dx.doi.org/10.3145/epi.2011.sep.09>
- Del Fresno, Miguel (2011b). *Netnografía*. Barcelona, Universitat Oberta de Catalunya.
- Del Fresno, M. (2012). "Comprendiendo los social media y mass media un modelo para el estudio de la comunicación interpersonal colectiva en tiempos de Internet", *Derecom*, 11, pp. 99-109
- Del Fresno, M. (2014). "Haciendo visible lo invisible: Visualización de la estructura de las relaciones en red en Twitter por medio del Análisis de Redes Sociales", *El profesional de la información*, 23(3), pp. 246-252 <http://dx.doi.org/10.3145/epi.2014.may.04>
- Del Fresno, M.; López, Antonio (2014). "Social work and netnography: the case of Spain and generic drugs". *Qualitative social work*, v. 13, n. 1, pp. 85-107. <http://dx.doi.org/10.1177/1473325013507736>

- Dorsey, J. (2012). "Twitter takes the pulse of the planet. It's the intersection of every media & médium", *Twitter*, November 15. Disponible en <http://mdelf.es/1b2uk> Consulta [30-09-14]
- Elmer, G. (2013). "Live research: *twittering* an election debate", *New media & society*, 15(1), pp. 18-30. <http://dx.doi.org/10.1177/1461444812457328>
- Easley, D.; Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge. Cambridge University Press. <http://dx.doi.org/10.1017/cbo9780511761942>
- Freeman, L.C. (2000). "Visualizing social networks", *Journal of Social Structure*, 1(1). Disponible en: <http://mdelf.es/12cp5> Consulta [30-09-14]
- Freeman, L.C. (2004). *The development of social network analysis: a study in the sociology of science*. Vancouver. Empirical Press.
- Geere, D. (2010). "It's not just you: 71 percent of tweets are ignored", *Wired*, November 11. Disponible en <http://mdelf.es/1b2up> Consulta [30-09-14]
- Gladwell, M. (2000). *The tipping point: How little things can make a big difference*. New York. Stoddart Publishing. <http://dx.doi.org/10.1080/15610950400010043>
- Golbeck, J. (2013). *Analyzing the Social Web*. Boston. Morgan Kaufmann. <http://dx.doi.org/10.1016/b978-0-12-405531-5.00016-x>
- Guandong X.; Lin L. (ed.). (2013). *Social media mining and social network analysis; emerging research*. Reference and Research Book News, 28(3). Portland: Ringgold Inc. <http://dx.doi.org/10.4018/978-1-4666-2806-9>
- Jinhui, T.; Meng, W.; Xian-Sheng, H.; Tat-Seng, Ch. (2012). "Social media mining and search", *Multimedia Tools and Applications*. (56), pp.1-7. <http://dx.doi.org/10.1007/s11042-011-0822-1>
- Lovejoy, K.; Waters, R.; Saxton, G. (2012). "Engaging stakeholders through Twitter: how nonprofit organizations are getting more out of 140 characters or less". *Public Relations Review*, 38, pp. 313-318. <http://dx.doi.org/10.1016/j.pubrev.2012.01.005>
- Mattelart, A. (2002). *Historia de la sociedad de la información*. Barcelona: Paidós.
- Mayer-Schönberger, V.; Cukier, K. (2014). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. Mariner Books. <http://dx.doi.org/10.2501/ija-33-1-181-183>

- Mitchell, W.J. (1995). *City of Bits: Space, Place, and the Infobahn*. Cambridge. MIT Press. <http://dx.doi.org/10.1017/s1359135500002797>
- Newman, M.E.J. (2010). *Networks*. Oxford. Oxford University Press.
- Schelling, T. (1978). *Micromotives and macrobehavior*. New York. Norton.
Disponible en <http://mdelf.es/1b4xz> Consulta [30-09-14]
- Sun, E.; Rosenn, I.; Marlow, C.; Lento, T. (2009). "Gesundheit! Modeling contagion through Facebook news feed", *Proceedings of the AAAI International Conference on Weblogs and Social Media*, pp. 146–153.
- Thomson, I. (2010). *Culture Wars and Enduring American Dilemmas*. Ann Arbor, MI: University of Michigan Press. <http://dx.doi.org/10.1086/660071>
- Virilio, P. (1991). *The aesthetics of disappearance*. New York: Semiotext(e).
Disponible en <http://mdelf.es/1b4y5> Consulta [30-09-14].
- Virilio, P. (1995). "Alerte dans le cyberspace!". *Le Monde diplomatique*, agosto.
Disponible en <http://mdelf.es/1b0xe> Consulta [30-09-14].
- Zafarani R.; Ali Abbasi, M.; Liu, H. (2014). *Social Media Mining*. Cambridge University Press. <http://dx.doi.org/10.1017/cbo9781139088510>
- Zimmerman, J. (2002). *Whose America? Culture Wars in the Public Schools*, Harvard University Press.