

ARTÍCULOS

PROPUESTA DE DISEÑO MUESTRAL PARA EL ANÁLISIS DE *Twitter* EN COMUNICACIÓN POLÍTICA

A sample design proposal for the analysis of *Twitter* in political communication

Salvador Percastre-Mendizábal, Carles Pont-Sorribes y Lluís Codina

Note: This article can be read in its original English version on:
<http://www.elprofesionaldelainformacion.com/contenidos/2017/jul/02.pdf>



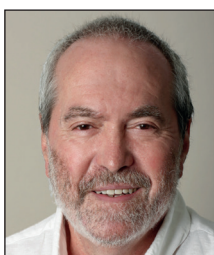
Salvador Percastre-Mendizábal es investigador doctoral en comunicación política en la *Universitat Pompeu Fabra (UPF)* y miembro del *Grup de Recerca en Comunicació Política, Mitjans i Democràcia*. Investigador doctoral adscrito al *ReSIC* de la *Université Libre de Bruxelles*. Investigador del proyecto competitivo *UPF* sobre *Comunicación 2.0 en emergencias* (financiado por la *Fundación BBVA*). Es miembro de la *European Communication Research and Education Association* y de la *Asociación Latinoamericana de Investigadores en Campañas Electorales*. Becario de *Conacyt-Cocit*.
<http://orcid.org/0000-0002-5947-5102>

salvador.percastre@upf.edu



Carles Pont-Sorribes es doctor en Comunicación Social, profesor-investigador en el *Departamento de Comunicación* de la *Universitat Pompeu Fabra (UPF)*. Investigador especializado en comunicación de riesgo y crisis, comunicación política, opinión pública y periodismo. Autor de artículos en revistas indexadas en el *Social Sciences Citation Index* y *Scopus*, entre otras, y de varios libros y capítulos de libro en editoriales especializadas. Es coordinador del grupo de investigación *Comunicación Política, Medios y Democracia (Polcom)* y subdirector de postgrado en el *Departamento de Comunicación* de la *UPF*.
<https://upf.edu/es/web/polcom>
<http://orcid.org/0000-0003-1135-4245>

carles.pont@upf.edu



Lluís Codina es profesor de la *Universitat Pompeu Fabra (UPF)* de Barcelona. Imparte docencia en la *Facultad de Comunicación*, en los Grados de Periodismo y de Comunicación Audiovisual. Es coordinador del *Máster Universitario en Comunicación Social (MUCS)* del *Departamento de Comunicación* de la *UPF*. Es miembro promotor del *Grupo de Investigación en Documentación Digital y Comunicación Interactiva (DigiDoc)*, coordinador de su *Seminario de Investigación* y coordinador del *Observatorio de Cibermedios*. Como miembro del *Grupo DigiDoc* forma parte de la *Unidad de Investigación en Periodismo*, grupo consolidado reconocido por la *Generalitat de Cataluña*.
<http://lluiscodina.com>
<http://observatoriocibermedios.upf.edu>
<http://orcid.org/0000-0001-7020-1631>

lluís.codina@upf.edu

Universidad Pompeu Fabra, Departamento de Comunicación
Roc Boronat, 138. 08018 Barcelona, España

Resumen

A partir del análisis de diversos ejemplos de discusiones en procesos de comunicación política en *Twitter*, se hace una propuesta metodológica para la obtención de una muestra de datos significativa, replicable y manejable de un universo de metadatos de *Twitter*, en investigaciones acerca de la comunicación política. La propuesta se basa en un nuevo modelo denominado *Top discussion indicator (TDI)* o Indicador de la máxima discusión. Su objetivo es ayudar a los investigadores a obtener un conjunto representativo de un universo que, aunque pueda superar cientos de miles de registros, incluya la mínima información que permita generalizar los resultados.

Palabras clave

Comunicación; Comunicación política; *Twitter*; Elecciones; Metodología; Métricas; *Inputs*; Interacciones.

Abstract

Based on an analysis of several sampled political discussions on *Twitter*, a new methodology is proposed that allows researchers to obtain a significant, replicable, and manageable data sample from a universe of *Twitter* metadata. The proposal is a new model called *Top discussion indicator (TDI)*. The aim of *TDI* is to assist researchers in obtaining a representative set of text from *Twitter* that includes the minimum amount of information needed to generalize the results.

Keywords

Communication; Political communication; *Twitter*; Elections; Methodology; Metrics; Inputs; Interactions.

Percastre-Mendizábal, Salvador; Pont-Sorribes, Carles; Codina, Lluís (2017). "A sample design proposal for the analysis of *Twitter* in political communication". *El profesional de la información*, v. 26, n. 4, pp. 579-588.

<https://doi.org/10.3145/epi.2017.jul.02>

1. Introducción

La misión de *Twitter*, según la propia plataforma de *micro-blogging* es

"dar a todos el poder de generar y compartir ideas e información al instante y sin obstáculos" (*Twitter*, 2016).

La compañía se define así como una plataforma de empoderamiento informativo global, es decir como un medio de la comunicación política digital (**Percastre-Mendizábal; Dorantes-Aguilar**, 2016).

El uso de *Twitter* como plataforma digital global de interacción social es indiscutible. Cuenta actualmente con 310 millones de usuarios activos mensualmente, registra 1.000 millones de visitas únicas mensuales a sitios web con acceso a la plataforma, se encuentra disponible en más de 40 idiomas y el 79 por ciento de usuarios provienen de fuera de Estados Unidos, según datos del mes de marzo de 2016 (*Twitter*, 2016).

Twitter se ha convertido en tendencia no sólo entre los usuarios de redes sociales; también se ha convertido en objeto de estudio de cada vez más investigadores en el mundo académico y empresarial. En el sector académico, se ha estudiado particularmente en los campos de (**Bruns; Stieglitz**, 2013):

- comunicación política
- comunicación de crisis
- comunicación de marca
- experiencias concretas de uso de *Twitter* como canal secundario o alterno (*backchannel*)
- relaciones interpersonales.

La mayoría de estudios de *Twitter* en comunicación política, se realizan desde una perspectiva metodológica cuantitativa y eminentemente positivista, con herramientas y modelos importados de las ciencias formales o desde la perspectiva de análisis estadístico, debido a los grandes volúmenes de datos puntuales y objetivos que es posible recolectar de la plataforma.

Uno de los problemas más habituales para los investigadores es la determinación de una muestra válida o la identificación segura de un conjunto de datos que permita realizar el análisis de una parte significativa del objeto de estudio.

Este artículo persigue reducir el volumen de datos a analizar y esto se puede hacer en base a determinar cronológicamente un punto en el tiempo, en el que se encuentre el mayor volumen de datos relacionados con un tema. A su vez, esto se puede realizar expresando el tema a través del uso de nodos semánticos de interconexión discursiva, como pueden ser las palabras clave (*keywords*), ya sea en forma de etiquetas (*hashtags*) o de menciones (*mentions*).

“*Twitter* se define como una plataforma de empoderamiento informativo global, es decir como un medio de comunicación política digital”

El principal objetivo de este trabajo, desde una perspectiva multidisciplinar de los estudios en comunicación, consiste en proponer un indicador válido que permita señalar un momento en el tiempo, en el desarrollo de un caso de estudio en comunicación política en *Twitter*, que configure un conjunto de datos que resulte óptimo y significativo para el análisis del caso. Este indicador representa una selección válida de *inputs* o entradas de datos en *Twitter* sobre un tema determinado en un momento concreto. Esta selección de información para su análisis, configura un conjunto de datos que reduce significativamente la gestión del universo de información recolectada en una base de datos.

2. Marco teórico. *Twitter* y comunicación política

En los estudios de *Twitter* en el campo de la comunicación política, destaca el de la comunicación en procesos electorales, particularmente lo relativo al estudio y análisis de elecciones y campañas electorales (**Jungherr; Schoen; Jürgens**, 2016; **Conway; Kenski; Wang**, 2015; **Larsson; Ihlen**, 2015). Además de los estudios sobre elecciones en *Twitter* en un contexto global, destacan estudios de caso que abordan temas específicos de la comunicación política electoral, principalmente en los siguientes aspectos:

- predicción electoral (**Burckhardt; Duch; Matsuo**, 2016; **Guerrero-Solé; Corominas-Murtra; López-González**, 2014);

- uso de *Twitter* en elecciones por parte de candidatos, partidos y ciudadanos (**Quevedo-Redondo; Portalés-Oliva; Berrocal-Gonzalo**, 2016; **Miller; Ko**, 2015);
- discusiones e información electorales (**López-Meri**, 2015; **López-García et al.**, 2015);
- comentarios simultáneos durante debates electorales (**Park et al.**, 2016; **Trilling**, 2015);
- impacto en los resultados electorales e influencia o persuasión en los electores (**Hosch-Dayican et al.**, 2016; **Vaccari et al.**, 2015);
- análisis de sentimiento en conversaciones electorales (**Himmelboim et al.**, 2014; **Stieglitz; Dang-Xuan**, 2012);
- viralidad de los mensajes en elecciones (**Congosto**, 2015; **Penney**, 2014);
- campañas negativas (**Ceron; D'Adda**, 2015; **Just et al.**, 2012).

Por otro lado, el estudio de fenómenos de comunicación política en esta red de microblogueo ha sido abordado también desde diversas disciplinas, que van desde la informática o las ciencias de la computación pasando por los enfoques disciplinares de la sociología, de la ciencia política o de las herramientas clásicas de la comunicación social como el análisis de discurso, el análisis desde la perspectiva teórica del encuadre mediático (*framing*) y el análisis de contenido, hasta los estudios lingüísticos formales o de la etnografía lingüística.

La mayoría de los estudios de *Twitter* sobre comunicación política se hacen desde una perspectiva metodológica cuantitativa y positivista utilizando modelos importados de las ciencias formales

Como demuestra un estudio del *Pew Research Center* de finales de 2014, hoy en día -al menos en EUA- la gente que se interesa en los asuntos públicos es susceptible de consumir contenido sobre política en *Twitter* por encima de otras redes sociales (**Gottfried**, 2014), por lo cual se puede afirmar que el acceso democrático a los medios sociales digitales ha impactado el *statu quo* de las jerarquías informativas tradicionales.

Los medios de comunicación preponderantes (*mainstream media*) han tenido que adaptarse a las nuevas lógicas informativas en el ámbito digital como demuestran algunos estudios empíricos (**Pont-Sorribes; Codina; Pedraza-Jiménez**, 2010). Con más razón ante la irrupción de los medios sociales, donde en muchos casos no significa tanto lo que alguien sea o lo que represente, sino lo que diga en sus redes sociales (*social networks*) o muestre a sus seguidores (*followers*), lo que comparta, publique, postee (*posting*) o lo que tweetee o retweetee.

Desde el surgimiento del uso de *Twitter* como plataforma de comunicación política, ha cambiado de manera fundamental la dinámica de las campañas electorales y de los procesos políticos, no sólo para que los candidatos y actores políticos puedan cumplir sus objetivos electorales o de gobierno, sino que ha cambiado la manera en cómo fluye la información entre ciudadanos, activistas y las propias élites políticas (**Ammann**, 2010).

3. Planteamiento metodológico

Para estudiar un fenómeno de comunicación política, resulta pertinente analizar los procesos de comunicación entre los 3 actores en un sistema social que esbozaba **Wolton** (1989), legitimados para expresarse acerca de los asuntos públicos y que en la actualidad pueden definirse y clasificarse en 3 grupos de actores:

- 1) Actores públicos: instituciones públicas, funcionarios públicos y políticos profesionales.
- 2) Actores mediáticos: medios de comunicación, periodistas y comunicadores.
- 3) Actores ciudadanos: pueden subdividirse en dos tipos:
 - grupos de interés específico, grupos de presión o ciudadanos organizados en colectivos y asociaciones con objetivos y causas concretas;
 - expresión colectiva de los ciudadanos a través de los instrumentos de medición de la opinión pública, que en la era de la hipermediación se expresan además en las plataformas de comunicación social digital, especialmente las redes sociales.

Los procesos de comunicación política digital, ya sea institucional, electoral o de gestión pública de una crisis o emergencia, comparten 3 elementos en común:

- por la propia naturaleza de la definición de un proceso de comunicación política, la participación de por lo menos algún actor de cada uno de los 3 grupos de actores que intervienen en la comunicación política;
- por la naturaleza de los eventos en los que tienen lugar los fenómenos de la comunicación política, tienen siempre un punto de inicio formal o informal pero que bien puede ser determinado en función de la naturaleza del evento como el período legal de duración de una campaña electoral, el período formal de atención institucional de un fenómeno de carácter político o la gestión oficial de una crisis; y
- dada la naturaleza discursiva de los eventos, existe siempre una curva natural que tiene un inicio, un desarrollo, un clímax y una conclusión.

A partir de la observación de diferentes casos de comunicación política en *Twitter*: una campaña electoral, una gestión de una emergencia y una crisis sanitaria de carácter público, encontramos la presencia de estas características de carácter replicable y equiparable. La obtención de los datos para la configuración de los universos de estudio se realizó a través de la extracción de *inputs* en *Twitter*: tweets y retweets a partir de etiquetas y menciones, de estos 3 casos seleccionados.

3.1. Métricas

La importancia de la determinación de los sistemas de medición a emplear resulta fundamental para la investigación, sin embargo, la determinación de las variables es un requisito indispensable para poder establecer el sistema de análisis con el que se han de realizar las métricas. Las variables determinadas no sólo atienden a la naturaleza del fenómeno a estudiar y a los objetivos de la investigación, sino también a la propia naturaleza del ecosistema en *Twitter*, es decir a los

elementos cuantitativos en esta plataforma y a las posibilidades de explotación de los mismos.

En función de las categorías conceptuales determinadas para una investigación, es necesario establecer las variables dependientes, las independientes y las de control, de acuerdo con las necesidades de cada modelo analítico.

Entre los elementos cuantificables a considerar se encuentran:

- seguidores;
- visitas;
- número de tweets y retweets;
- perfiles o cuentas más influyentes;
- perfiles que más siguen a otros perfiles;
- palabras clave: pueden ser etiquetas (*hashtags*) o menciones (*mentions*);
- bioperfiles o elementos descriptivos del perfil de *Twitter*;
- geolocalización de los perfiles, palabras clave o discusiones;
- principales temas del momento (*trending topics*);
- comentarios positivos, negativos o neutros;
- tweets como respuestas a discusiones fuera de línea (*offline*) como en el caso de las discusiones en *Twitter* sobre los debates televisivos.

El análisis de las siguientes variables resulta indispensable para comprender los procesos de comunicación política en *Twitter* y sirven de base para el establecimiento del modelo de análisis:

A. Métricas básicas

Son aquellos elementos eminentemente cuantificables que permiten describir cuantitativamente la importancia, relevancia, presencia o jerarquía de los perfiles de *Twitter*. Esto permite hacer una primera valoración del peso de la audiencia digital, es decir el número potencial de electores, simpatizantes o seguidores de un perfil determinado, en el caso de los actores políticos; relevancia también política pero eminentemente informativa, de medios de comunicación, periodistas y comunicadores; y también de perfiles ciudadanos, tweeteros o *twitter-activistas*.

a) Número de seguidores: perfiles de *Twitter* que siguen a la cuenta (perfil) objeto de análisis o que forma parte del universo de estudio, en el preciso momento de la recolección de datos. El número de seguidores puede variar considerablemente en un período corto de tiempo, sobre todo en perfiles de candidatos y partidos, ya que uno de los objetivos de la campaña en redes sociales digitales es justo obtener el mayor número de seguidores del área geográfica en la que se concentra la campaña, el proceso electoral o la demarcación de la crisis institucional.

b) Cuentas a las que se sigue (*following*): número de perfiles que siguen las cuentas que forman parte del estudio, exactamente en el momento en que se realiza la recogida de los datos.

c) Retweets recibidos: todos los retweets que el perfil ha recibido del total de las publicaciones relacionadas con el caso de estudio, ya sea delimitadas temporalmente o por palabras clave: etiquetas o menciones.

d) "Me gusta" (*like*): número total de mensajes marcados por otros usuarios como "me gusta" -anteriormente denominado "favorito" (*favorite*)-, de una cuenta o perfil que for-

ma parte del análisis, que se integra al universo de estudio y del cual se desprende el volumen de datos a analizar.

B. Inputs

Se describe el conjunto de entradas en la plataforma, que los usuarios realizan a través de su cuenta de *Twitter*, es decir: mensajes o información que los usuarios introducen redactando los tweets, copiando y pegando información en forma de texto o de contenido audiovisual o interactivo, que constituye el contenido del mensaje. Generalmente se realizan en forma manual, aunque también es posible realizarlos en forma automatizada a través de un programa informático o de una aplicación robot.

a) Tweets propios: mensajes que emite en forma original el usuario de la cuenta; el contenido puede estar en caracteres alfanuméricos, vídeos, audios, podcasts, gifs, fotografías, ilustraciones, gráficos, afiches digitales, memes, hipervínculos, menciones a otras cuentas de *Twitter*, emoticonos, etc.

b) Retweets realizados: mensajes de *Twitter* realizados por el mismo usuario o por otro distinto, que son reenviados a través de la propia cuenta; es decir, es un tweet que puede ser propio o no, pero que es tuiteado nuevamente.

c) Total de *inputs*: corresponde al cómputo o recopilación de todas las entradas de mensajes e información contenida en los propios tweets y/o las propias unidades de *Twitter*, es decir los tweets.

C. Interacciones

Destacan aquí los elementos de *Twitter* que por su naturaleza intrínseca funcionan como recursos de interacción para establecer relaciones entre usuarios ya sea directamente a través de las menciones expresas del identificador de usuario o de nodos de conexión hipertextual entre usuarios, acerca de temas, como en el caso de las etiquetas.

a) Etiquetas (*hashtags*): palabras, frases o conjuntos de caracteres que expresan un concepto o idea o que mencionan un asunto de interés relevante en las discusiones en la tuitesfera, se escriben inmediatamente precedidos del símbolo de almohadilla y en el caso de contener más de una palabra, se suprimen los espacios.

b) Menciones a otros usuarios: cuando un perfil menciona a otro usuario en un mensaje de *Twitter*, ya sea en forma única o generalmente como parte del contenido de un mensaje, escribe el identificador del nombre de usuario o "ID de usuario", es decir la denominación oficial de la cuenta precedida del símbolo de arroba; a esto se le llama una mención de usuario.

c) Enlaces compartidos: enlaces informáticos en forma de hipervínculos, que contienen la ruta de la dirección electrónica para acceder a contenido fuera o dentro de la propia plataforma, aunque generalmente se comparte contenido externo como webs de otras redes sociales.

d) Total de interacciones: cómputo del total de las interacciones que se determinen para el estudio, ya sean las propias unidades de interacción como las menciones o los hiperenlaces, ya sean los propios mensajes que contienen alguna interacción, siempre y cuando se establezca y justifique el parámetro para realizar el cómputo de unas u otras interacciones.

3.2. Modelo de selección de los datos: *Top discussion indicator (TDI)*

Existen múltiples formas de abordaje metodológico de estudios de caso de comunicación política en *Twitter*. Sin embargo la selección de las muestras es aleatoria y no atiende a criterios únicos. **Jungherr, Schoen y Jürgens** (2016) usaron para su estudio 6.677.795 tweets previos a una campaña electoral, publicados por 1.248.667 usuarios.

Burckhardt, Duch y Matsuo (2016) recolectaron 25.000.000 de tweets con los nombres de los seis partidos más grandes del Reino Unido y de sus líderes, de 5 meses y medio previos a la campaña electoral; sin embargo, para el análisis se utilizó únicamente una muestra aleatoria de 8.000.000 de tweets. **Hosch-Dayican et al.** (2016) analizaron 368.855 tweets recolectados durante dos semanas y media, clasificados por etiquetas, realizando un filtro primero por idioma (neerlandés) y luego por ubicación, dejando fuera los tweets irrelevantes generados por un algoritmo propio de selección y dejando a los directamente relacionados con la elección parlamentaria, los partidos y los candidatos.

López-Meri (2015) analiza 500 tweets aleatoriamente en un rango de 4 días correspondientes al fin de semana previo a unas elecciones catalanas.

Miller y Ko (2015) analizan los mensajes de 50 miembros electos al parlamento de Kuwait durante una campaña electoral.

Ceron y D'Adda (2015) analizan los mensajes de los 8 partidos italianos más grandes y de sus líderes, durante la campaña electoral, un total de 15.053 tweets.

Congosto (2015) estudia los mensajes publicados por quien encabezó la lista electoral o por la cuenta de cada uno de los partidos españoles, así como las menciones recibidas y los mensajes relacionados con el debate electoral determinado. Aunque obtiene una muestra de 1.552.282 mensajes, sólo clasifica 536.588, un 35,18% del total.

Aun cuando pareciera que no existen límites a la generación de información en redes sociales digitales, existen márgenes “naturales” que permiten delimitar de una forma razonable los universos de datos de fenómenos de comunicación política en *Twitter*. Es el caso por ejemplo de los mensajes emitidos por perfiles determinados geográficamente, por temas definidos con palabras clave (*keywords*) o por la temporalidad de los mensajes.

A partir de varios casos de análisis, se propone un indicador cronológico para determinar un momento de recogida de datos que resulte no sólo más significativo y reduzca en varios órdenes de magnitud el volumen de los datos, sino que plausiblemente evite “ruido” e impida considerar momentos poco significativos para la recogida de datos.

Con este modelo se plantea obtener un conjunto de usuarios y conversaciones en *Twitter* que represente el mayor volumen de las discusiones en el tiempo, sobre un tema de comunicación política ya sea electoral, de gestión institucional o de crisis pública, a efecto de obtener un momento en el tiempo a lo largo del desarrollo de la crisis, en el que se haya efectuado la mayor cantidad de *inputs* sobre un tema determinado y en el que convergen al mismo tiempo la ma-

yor cantidad de usuarios en un momento concreto, que se expresan e interactúan sobre temas definidos a través de palabras clave.

Independientemente del caso de estudio y del volumen de información que se genere en una base de metadatos de *Twitter*, es posible obtener un indicador para obtener un volumen plausible de datos de un universo determinado, aquí denominado *Top discussion indicator (TDI)*, de tal manera que se podrá encontrar y señalar un momento preciso, que bien puede ser un día, una franja horaria o una hora en concreto, en que se concentra el mayor volumen de discusiones de todas las palabras clave, ya sean menciones o etiquetas, en un mismo período de tiempo.

Lo anterior permite deducir de forma importante el mayor número de discusiones de todos los actores que se expresaron sobre los temas y el mayor volumen de impactos potenciales debido a la cantidad de mensajes en el momento más álgido de las discusiones.

Esto permite analizar en forma eficiente una porción de discusiones y perfiles que participaron en el momento más significativo y probablemente más relevante en toda la curva cronológica del desarrollo del propio fenómeno de comunicación política.

Ante elevados volúmenes de información como en los casos analizados que rebasan los cientos de miles de registros, resulta útil poder tomar ese momento preciso en el que se lleva a cabo el mayor volumen de discusiones y en el que participa el mayor número de actores, ya que como conjunto de datos permite obtener un momento relevante y significativo del fenómeno a estudiar, además de reducir las posibilidades de ruido con muestras demasiado amplias y poco o nada fundamentadas.

4. Casos de estudio

Se analizan tres casos de comunicación política en los que se generó una actividad significativa en *Twitter*:

- elecciones federales mexicanas de 2015;
- accidente del vuelo 9525 de *Germanwings*;
- crisis del ébola en España en 2014.

4.1. Elecciones mexicanas

El primer caso de análisis es la elección parlamentaria federal en México en 2015. El 7 de junio de ese año se llevó a cabo en las 32 Entidades Federativas de ese país la elección para renovar a los 500 miembros de la *Cámara de Diputados del Congreso de la Unión*, en una de las 3 jornadas electorales con más alta participación en la historia del país norteamericano.

El período legal de campaña electoral comprendió del 5 de abril al 4 de junio del mismo año, es decir, tuvo una duración de 61 días. Durante este período se recogió un universo de datos de todas las discusiones en *Twitter* en torno a la etiqueta *#elecciones2015*, que realizaron la función de lo que aquí se denomina: nodos semánticos digitales de interconexión discursiva, a efecto de obtener las conversaciones relacionadas con las elecciones de todos los usuarios que emitieron alguna entrada en *Twitter*, en forma de tweet convencional (original), retweet directo o retweet modifi-

cado, conocido como MT (*modified tweet*).

Una vez aplicados los filtros para discriminar los usuarios por ubicación geográfica manifiesta en la descripción o biografía del perfil (*bio*) y por idioma, se obtuvo un volumen de datos con más de 250.000 registros, los cuales cronológicamente ordenados y graficados, arrojan el resultado que se muestra en el gráfico 1. Se observa que existe un único momento, un día específico de la campaña durante el cual se registró el mayor volumen de discusión. Fue donde mayor número de entradas se generaron en torno a la palabra clave analizada y, por lo tanto, el día de mayor masa discursiva generada durante el caso de estudio.



Gráfico 1. Evolución cronológica del número de tweets con la etiqueta #elecciones2015 durante la campaña legal federal en México en 2015

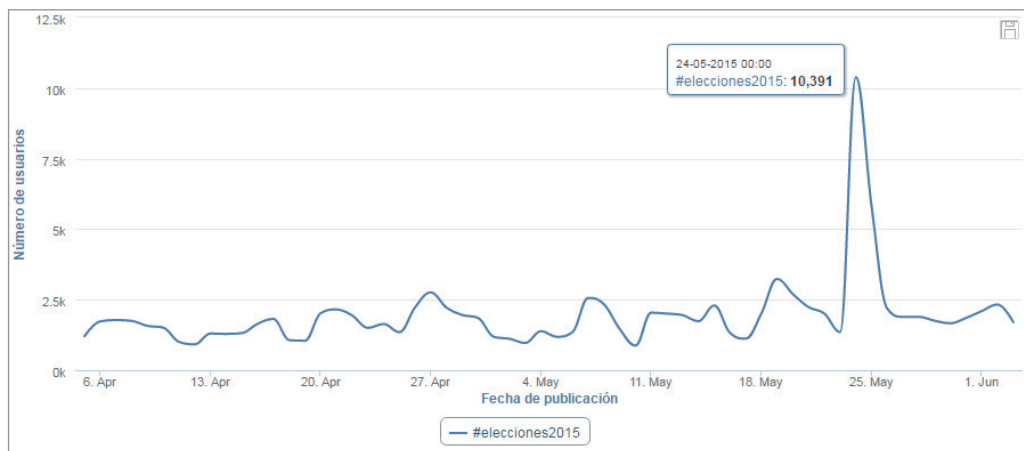


Gráfico 2. Evolución cronológica del número de usuarios activos interactuando con la etiqueta #elecciones2015, durante la campaña legal federal en México en el año 2015

De 250.000 registros del universo se obtiene un volumen de datos correspondiente a 15.964 registros en el día de mayor discusión, en este caso el 24/05/2015.

Aplicando el *TDI* de este universo de 250.000 registros, 684 tuvieron por lo menos 50 retweets y de éstos, 548 obtuvieron 10 o más “me gusta”, lo que representa el 0,21% del total. En este caso, la muestra a analizar se reduce sustancialmente de un cuarto de millón de registros a poco más de quinientos.

El día en que se generó el mayor número de *inputs* en *Twitter*, también se registró el mayor número de usuarios activos interactuando en torno a la palabra clave en forma de etiqueta #elecciones2015, como lo demuestra el gráfico 2; es decir 10.391 usuarios.

4.2. Accidente de Germanwings

El segundo caso de análisis fue la emergencia por el accidente del vuelo 9525 de *Germanwings*, cuando el avión de esta compañía alemana, que despegó del aeropuerto de *Barcelona-El Prat* con destino a *Düsseldorf*, se estrelló en los Alpes franceses de Provenza el 24 de marzo de 2015 alrededor de las 10:30 horas, con 150 personas a bordo que murieron en el acto. Las investigaciones concluyeron que el accidente fue causado intencionadamente por el copiloto del avión.

En el gráfico 3 se observa claramente el mismo fenómeno

de discusiones hiperconcentradas en un punto en el tiempo, durante el período de desarrollo de las discusiones sobre la crisis mencionada, que se inicia el 24 de marzo cuando se dan las primeras informaciones de la desaparición de vuelo en el radar y culmina el 22 de julio, cuando se dieron las últimas informaciones sobre el comunicado oficial de la aerolínea anunciando las indemnizaciones a los deudos.

Durante el período de recolección de información, se generó un universo de datos con 235.829 *inputs* de las discusiones en *Twitter* con las palabras clave:

- #Germanwings, referente al nombre de la compañía;
- #4U9525, número oficial del vuelo;
- Dusseldorf, ciudad destino del vuelo siniestrado;
- Estrop, pico Tête de l’Estrop en cuyas cercanías se estrelló el aparato.

Se observa que todas las etiquetas y menciones tuvieron el mismo momento de mayor volumen de información al respecto, el día 24/03/2015. Se recogieron 103.909 *inputs* con las etiquetas y menciones más usadas durante la crisis. En este caso, del universo total de 235.829 registros, filtrando los *inputs* más destacados, únicamente 1.273 tuvieron 50 o más retweets y de éstos, los que tuvieron 10 o más “me gusta” fueron 481, es decir el 0,35% del total inicial.

En el gráfico 4 observamos que el número de usuarios coincide con el día en el que tuvo lugar el mayor volumen de

participación de los usuarios. Cuanto mayor es el número de menciones, mayor es el número de usuarios que interactúan con esas menciones como en el caso de la etiqueta *#Germanwings*, que fue la más empleada por los usuarios. Mientras que la suma total de usuarios interactuando con las 4 etiquetas es de 66.784.

4.3. Crisis del ébola en España

El tercer caso de análisis se refiere a la llamada crisis del ébola en España. Esta crisis sanitaria se inició con el primer contagio por el virus del ébola en territorio español, el de Teresa Romero, una auxiliar de enfermería en el *Hospital Carlos III de Madrid*, quien -en primera instancia- asistió al religioso Miguel Pajares, primer español infectado con el virus, tras ser repatriado desde Liberia, a raíz del brote de esa enfermedad en África Occidental en 2014. Se informó del caso el 6 de octubre de ese año, convirtiéndose en el primero que se originaba en suelo español y también europeo.

Se recolectaron 1.880.750 registros de las discusiones en torno a las etiquetas:

- #ebolaenespana
- #ebolaenespaña
- #salvemosalexcalibur
- #anamatodimision
- #JavierRodriguezDimision
- #teresaromero
- #TodosSomosTeresa
- #vamosamorirtodos

Los registros se recolectaron en el período comprendido entre el 25 de junio de 2014, fecha en la que se activa por primera vez el protocolo de emergencia en Europa por lo que fuera una falsa alarma de un caso de contagio de ébola en Valencia, y el 2 de diciembre de 2014, fecha de la declaración oficial del fin de la emergencia por el ébola en España por parte de la *Organización Mundial de la Salud*.

En este otro caso observamos cómo de un total de 1.880.750 registros de *Twitter*, se decantaron siguiendo el sistema planteado aquellos de 50 o más retweets, resultando 3.360



Gráfico 3. Evolución cronológica del número de tweets con las palabras clave sobre la crisis por el accidente del vuelo 9525 de *Germanwings*

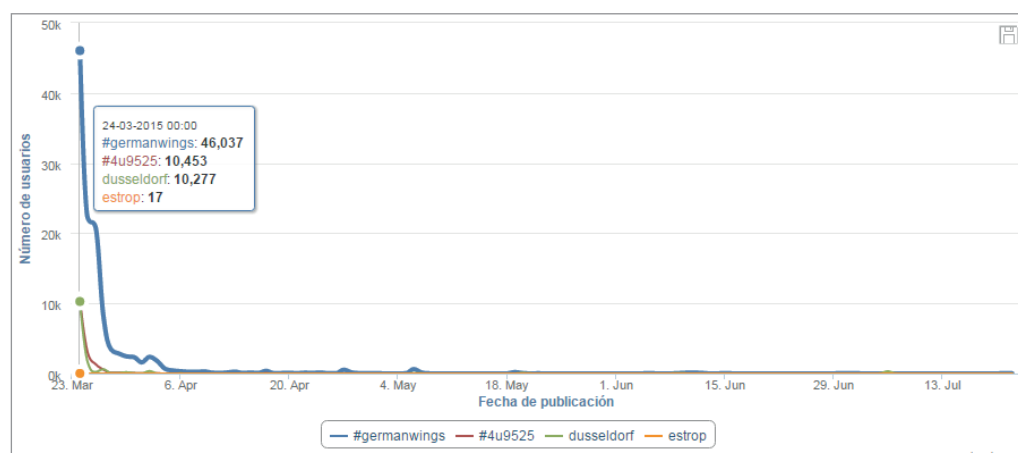


Gráfico 4. Evolución cronológica por número de usuarios de los *inputs* realizados con las palabras clave sobre la crisis por el accidente del vuelo 9525 de *Germanwings*

registros, de los cuales 1.113 presentaron 10 o más “me gusta”, el 0,05%.

Como se observa en los gráficos 5 y 6, se registra el mismo comportamiento de las discusiones en los casos anteriores, concentrándose en un día determinado del período de análisis, el 08/10/2014; presentándose 358.290 registros. Se replica tanto el mayor número de tweets enviados en torno a las etiquetas más relevantes y también se presenta el mayor número de usuarios participando en torno a estos mismos temas.

Para determinar el *TDI* se propone el establecimiento de dos criterios: marco temporal y factor de viralización.

El primero consiste en establecer el momento preciso de la mayor actividad. Dependiendo del caso de estudio y del volumen de *inputs*, puede ser un día, una franja horaria o una hora exacta a fin de localizar el momento de mayor auge de conversación.

Una vez aplicada esta primera decantación vinculada al marco temporal y sumando los resultados de todos los *inputs* originales publicados, es decir, los tweets originales o generados directamente por el propio usuario de *Twitter*, excluyendo los retweets, se establece un criterio de relevancia para identificar los tweets de mayor impacto de la selección.

Para ello, se consideran *inputs* originales de alto grado de impacto y por tanto relevantes, los *inputs* que de entre

todos los usuarios hayan tenido preferencia efectiva y positiva, es decir aquellos mensajes que como mínimo cuenten con 50 retweets; esto representa el 50% de los últimos usuarios que *Twitter* muestra que retweetearon un tweet público -como señala el propio sitio web de la plataforma- y que por la experiencia de la observación de los procesos de viralización de los mensajes en *Twitter*, se considera una cantidad suficiente de retweets para ser catalogado como contenido relevante.

De estos últimos mensajes, los que hayan registrado por lo menos 10 “me gusta”, es una cifra aleatoria de discriminación que sirve de parámetro para la obtención de un segundo filtro de calidad para seleccionar un menor número de *inputs*, que por sí mismos tuvieron un interés cualitativo superior por parte de los perfiles de usuario.

Este parámetro se determina como un criterio con valores numéricos aleatorios, que sirve de mecanismo de discriminación significativo para seleccionar los mensajes más relevantes. El hecho de otorgar mayor valor numérico al retweet que a los “me gusta” es consecuencia de dos cuestiones:

- en la lógica interna de la línea del tiempo del perfil de *Twitter* (*timeline*), un retweet otorga mayor visibilidad a los mensajes que un “me gusta”, por lo que su valor como herramienta de viralización del contenido es más alto;
- como ha demostrado el propio análisis, la norma general es que los tweets obtengan un mayor número de retweets que de “me gusta”, por lo que resulta necesario otorgarle un valor numérico más elevado.

El microanálisis en fases ha consistido en identificar cuáles de estas fases aparecen en el contexto del *Top discussion indicator*, es decir, en el marco de una temporalidad reducida, de apenas un día en cada caso, y en mensajes de alta viralización. Esto permite saber qué fase fue la más destacada en el punto álgido de la conversación, es decir, el día que más se publicó sobre el propio caso de estudio y entre los mensajes que más impactos (retweets y “me gusta”) recibieron.

La virtud de este indicador es que permite realizar análisis cualitativo al tener una porción de datos plausible y singu-

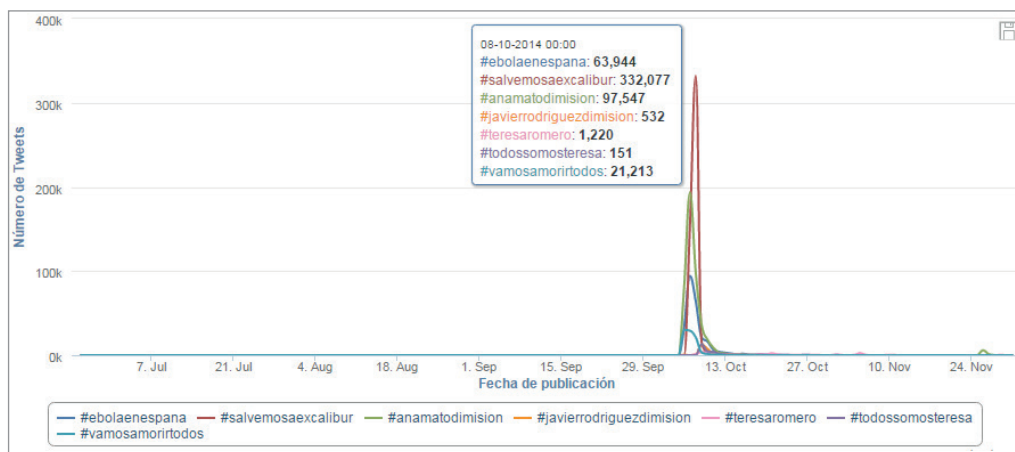


Gráfico 5. Evolución cronológica del número de tweets con las palabras clave sobre la crisis del ébola en España

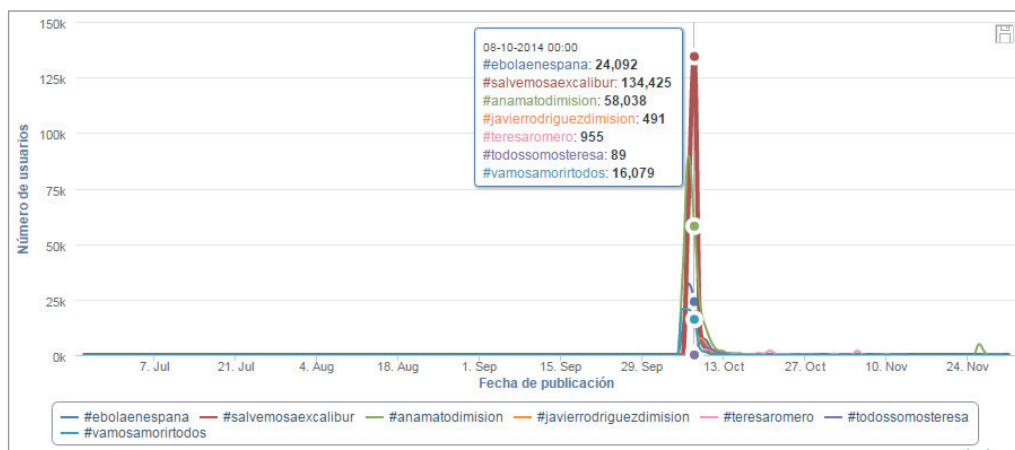


Gráfico 6. Evolución cronológica por número de usuarios de los *inputs* realizados con las palabras clave sobre la crisis del ébola en España

larmente más reducida, con la cual poder realizar análisis etnográfico, semántico o discursivo, como se observa en el análisis de los 3 casos anteriores.

5. Discusión y conclusiones

La determinación del conjunto significativo de registros a analizar a partir de un universo de metadatos de *Twitter*, podría expresarse algebraicamente como sigue:

$$cTDI = (U-tx)-(m)-(s)$$

Donde:

- *cTDI* o conjunto *TDI*: representa el conjunto de datos a analizar del universo;
- *U*: universo de metadatos obtenidos;
- *tx*: datos extraídos de la franja precisa en el tiempo y determinada para cada caso, que pueden ser días, horas, minutos o segundos;
- *m*: todos los *inputs* de menor impacto, es decir aquellos con menos de 50 retweets como factor de viralización;
- *s*: *inputs* señalados con menos de 10 “me gusta”.

Debido al volumen de información y datos que es posible obtener de *Twitter*, desde la investigación en comunicación y desde la perspectiva del análisis cualitativo existen dificultades en el tratamiento y selección de la información a analizar. Eso sucede no sólo por el volumen de los propios datos, sino por la propia naturaleza de almacenamiento de la información, que en muchas ocasiones resulta complejo

administrar, clasificar y analizar con herramientas tradicionales de estudio, incluso las informáticas.

Si bien el diseño muestral a partir del *TDI* permite reducir las muestras de tweets y obtener un conjunto de datos más abarcables para el análisis, es posible realizar el procesamiento de datos empleando metodologías cualitativas o cuantitativas. Aunque al ser un indicador que permite refinar un alto volumen de registros de *Twitter*, su mayor utilidad y valor será para investigaciones con enfoque cualitativo.

En cuanto a la explotación de los datos, la mayor parte de los estudios se realizan con métodos cuantitativos, por ejemplo, análisis de contenido, análisis de sentimiento, análisis de redes o incluso con un alcance descriptivo desde la estadística analítica. Ello ha dificultado no sólo la gestión del volumen de la información, sino la determinación de los modelos de análisis a emplear, las herramientas a usar para la recolección de los datos e incluso la especialización de los investigadores para conocer el manejo de estas nuevas medidas de información. En función del método y de los objetivos de la investigación, los investigadores filtran metodológicamente su universo de estudio para obtener muestras que permitan un manejo eficiente de la información y adaptado a las herramientas tanto metodológicas como informáticas para su análisis.

Es probable que este indicador muestral resulte útil y efectivo para aplicarse en muestras de otros fenómenos de la comunicación o de las ciencias sociales; sin embargo, con la evidencia disponible, aún no es posible demostrarlo. Hasta ahora las pruebas indican que funciona plausiblemente en comunicación política. A investigadores de otras disciplinas les tocaría decir -y demostrar- si funciona también en ellas. Probablemente resulte así, sin embargo, se requiere de mayor evidencia empírica.

La originalidad del *TDI* resulta del hecho de que aún no se han encontrado criterios replicables y universales para la proporcionalidad de la determinación de muestras bajo principios metodológicos razonables y criterios de representatividad, que resulten plausibles. Los investigadores de estos fenómenos se encuentran aún en un proceso de búsqueda del método más eficiente para el análisis de este nuevo medio, como sucedió en los años 60 de la década pasada con el surgimiento de la televisión.

Existe no sólo una amplia gama de modelos y técnicas de abordaje metodológico así como de herramientas de análisis, sino que resulta significativo que para realizar estudios de caso de temáticas vinculadas y fenómenos muy similares epistemológicamente, se use una variación -en algunos casos diametralmente distante- de métodos de ponderación, de determinación de variables y establecimiento de categorías, así como en la configuración de bases de datos y en las herramientas informáticas para la recolección de la información y la determinación de las muestras.

Por ello el *TDI* se plantea como un recurso metodológico para la investigación, que puede determinar un *temporis momentum* que permite obtener un conjunto de datos refinado, significativo y relevante, de fenómenos cronológicos de discusión en comunicación política, evitando analizar todo el universo de datos recolectado en torno a uno o varios temas en *Twitter*.

Sin embargo, junto a la probada utilidad metodológica del *TDI*, resulta pertinente señalar dos carencias:

- las pruebas empíricas aún no permiten universalizar su aplicación en otras disciplinas concomitantes de la comunicación política;
- aun cuando la muestra final obtenida es relevante, plausible y significativa, no es posible determinar una correlación estadística que permita encontrar una relación proporcional de la muestra con respecto al universo; sin embargo, al menos hasta el momento, esto no se ha podido determinar inequívocamente con modelos estadísticos, como en el caso de los estudios de opinión pública.

Por tanto, esta propuesta pretende abonar en la discusión y propiciar pautas de abordaje metodológico que ayuden a estandarizar criterios de indagación académica en esta materia, y que posibiliten a nuevas investigaciones contrastar empíricamente las formulaciones aquí planteadas.

Nota

Este trabajo se desprende del proyecto científico *Comunicar en situaciones de emergencia. Herramientas 2.0 y nuevos protocolos en la gestión eficiente de la comunicación de emergencia*, financiado por la Fundación BBVA, y es parte del proyecto *Creación y contenido interactivo en la comunicación de información audiovisual: audiencias, diseño, sistemas y formatos* (CSO2015-64955-C4-2-R, Mineco/Feder), Ministerio de Economía y Competitividad (España).

6. Bibliografía

- Ammann, Sky** (2010). *A political campaign message in 140 characters or less: The use of Twitter by U.S. Senate candidates in 2010*. SSRN Papers. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1725477
- Bruns, Axel; Stieglitz, Stefan** (2013). "Towards more systematic Twitter analysis: metrics for tweeting activities". *International journal of social research methodology*, v. 16, n. 2, pp. 91-108. <https://eprints.qut.edu.au/58132> <https://doi.org/10.1080/13645579.2012.756095>
- Burckhardt, Philipp; Duch, Raymond; Matsuo, Akitaka** (2016). "Tweet as a tool for election forecast: UK 2015. General election as an example". En: *Third annual meeting of the Asian Political Methodology Society in Beijing*. <https://goo.gl/8wVxzB>
- Ceron, Andrea; D'Adda, Giovanna** (2015). "E-campaigning on Twitter: The effectiveness of distributive promises and negative campaign in the 2013 Italian election". *New media & society*, v. 18, n. 9, pp. 1935-1955. <https://goo.gl/xZEDnC> <https://doi.org/10.1177/1461444815571915>
- Congosto, María-Luz** (2015). "Elecciones europeas 2014: viralidad de los mensajes en Twitter". *Redes. Revista hispana para el análisis de redes sociales*, v. 26, n. 1, pp. 23-52. <https://doi.org/10.5565/rev/redes.529>
- Conway, Bethany; Kenski, Kate; Wang, Di** (2015). "The rise of Twitter in the political campaign: Searching for intermedia agenda-setting effects in the presidential primary". *Journal of computer-mediated communication*, v. 20, n. 4, pp. 363-380. <https://goo.gl/3j2DKG>

<https://doi.org/10.1111/jcc4.12124>

Gottfried, Jeffrey (2014). "Facebook and Twitter as political forums: Two different dynamics". *Pew Research Center*, 12 November.

<http://www.pewresearch.org/fact-tank/2014/11/12/facebook-and-Twitter-as-political-forums-two-different-dynamics/>

Guerrero-Solé, Frederic; Corominas-Murtra, Bernat; López-González, Hibai (2014). "Pacts with Twitter. Predicting voters' indecision and preferences for coalitions in multiparty systems". *Information, communication & society*, v. 17, n. 10, pp. 1280-1297.

<https://goo.gl/XvWq9G>

<https://doi.org/10.1080/1369118X.2014.920040>

Himelboim, Itai; Sweetser, Kaye; Tinkham, Spencer; Cameron, Kristen; Danelo, Matthew; West, Kate (2014). "Valence-based homophily on Twitter: Network analysis of emotions and political talk in the 2012 presidential election". *New media & society*, v. 18, n. 7, pp. 1382-1400.

<https://doi.org/10.1177/1461444814555096>

Hosch-Dayican, Bengu; Amrit, Chintan; Aarts, Kees; Dassen, Adrie (2016). "How do online citizens persuade fellow voters? Using Twitter during the 2012 Dutch parliamentary election campaign". *Social science computer review*, v. 34, n. 2, pp. 135-152.

<https://doi.org/10.1177/0894439314558200>

Jungherr, Andreas; Schoen, Harald; Jürgens, Pascal (2016). "The mediation of politics through Twitter: An analysis of messages posted during the campaign for the German federal election 2013". *Journal of computer-mediated communication*, v. 21, n. 1, pp. 50-68.

<https://goo.gl/zBw3dz>

<https://doi.org/10.1111/jcc4.12143>

Just, Marion; Metaxas, Panagiotis; Mustafaraj, Eni; Crigler, Ann (2012). *It's trending on Twitter: An analysis of the Twitter manipulations in the Massachusetts 2010 special Senate election*.

<http://www.researchgate.net/publication/256026592>

Larsson, Anders; Ihlen, Øyvind (2015). "Birds of a feather flock together? Party leaders on Twitter during the 2013 Norwegian elections". *European journal of communication*, v. 30, n. 6, pp. 666-681.

<https://goo.gl/N9YpYF>

<https://doi.org/10.1177/0267323115595525>

López-García, Guillermo; Gámir-Ríos, José-Vicente; García-Ull, Francisco-José; Llorca-Abad, Germán; Cano-Orón, Lorena; González-Esteban, José-Luis (2015). "El debate sobre Europa en Twitter. Discursos y estrategias de los candidatos de las elecciones al Parlamento Europeo de 2014 en España". *Revista de estudios políticos* (nueva época), n. 170, pp. 213-246.

<https://doi.org/10.18042/cepc/rep.170.07>

López-Meri, Amparo (2015). "Redes sociales y campañas electorales: Twitter como fuente informativa en las elecciones catalanas del #25N". *Comunicació: Revista de recerca i d'anàlisi*, v. 32, n. 2, pp. 115-137.

<https://goo.gl/duUNdN>

Miller, Noah; Ko, Rosa (2015). "Studying political microblogging: Parliamentary candidates on Twitter during the

February 2012 election in Kuwait". *International journal of communication*, v. 9, pp. 2933-2953.

<http://ijoc.org/index.php/ijoc/article/view/1910/1464>

Park, Se-Jung; Park, Ji-Young; Lim, Yon-Soo; Park, Han-Woo (2016). "Expanding the presidential debate by tweeting: The 2012 presidential election debate in South Korea". *Telematics and informatics*, v. 33, n. 2, pp. 557-569.

<https://goo.gl/nZHJCP>

<https://doi.org/10.1016/j.tele.2015.08.004>

Penney, Joel (2014). "Motivations for participating in 'viral politics': A qualitative case study of Twitter users and the 2012 US presidential election". *Convergence: The international journal of research into new media technologies*, v. 22, n. 1, pp. 71-87.

<https://goo.gl/BiaWgU>

<https://doi.org/10.1177/1354856514532074>

Percastre-Mendizábal, Salvador; Dorantes-Aguilar, Gerardo-Luis (2016). "Comunicación política electoral en la era digital. La campaña electoral de 2015 en la ciudad de México". *Vivat academia. Revista de comunicación*, diciembre, n. 137, pp. 36-56.

<https://goo.gl/y4uRRP>

Pont-Sorribes, Carles; Codina, Lluís; Pedraza-Jiménez, Rafael (2009). "Comunicación de riesgo y sistemas de información en la Web: cinco modelos". *El profesional de la información*, v. 18, n. 4, pp. 389-397.

<https://doi.org/10.3145/epi.2009.jul.05>

Quevedo-Redondo, Raquel; Portalés-Oliva, Marta; Berrocal-Gonzalo, Salomé (2016). "The image use on Twitter during the 2015 municipal election campaign in Spain". *Revista latina de comunicación social*, v. 71, pp. 85-107.

<https://doi.org/10.4185/RLCS-2016-1085en>

Stieglitz, Stefan; Dang-Xuan, Linh (2012). "Political communication and influence through microblogging – An empirical analysis of sentiment in Twitter messages and retweet behavior". En: *IEEE computer society*, pp. 3500-3509.

<https://goo.gl/PswMJK>

<https://doi.org/10.1109/HICSS.2012.476>

Trilling, Damian (2015). "Two different debates? Investigating the relationship between a political debate on TV and simultaneous comments on Twitter". *Social science computer review*, v. 33, n. 3, pp. 259-276.

<https://goo.gl/ve2AMe>

<https://doi.org/10.1177/0894439314537886>

Twitter (2016). About.

<https://about.twitter.com/company>

Vaccari, Cristian; Valeriani, Augusto; Barberá, Pablo; Bonneau, Rich; Jost, John; Nagler, Jonathan; Tucker, Joshua (2015). "Political expression and action on social media: Exploring the relationship between lower- and higher-threshold political activities among Twitter users in Italy". *Journal of computer-mediated communication*, v. 20, n. 2, pp. 221-239.

<https://doi.org/10.1111/jcc4.12108>

Wolton, Dominique (1989). "La communication politique: construction d'un modèle". *Hermès*, n. 4, pp. 27-42.

<https://doi.org/10.4267/2042/15353>