

*Análisis de la dinámica, la estructura y el contenido de los mensajes de Twitter: violencia sexual en #Cuéntalo **

Dynamic, structural and content analysis of Twitter: sexual violence in #Cuéntalo

MODESTO ESCOBAR MERCADO

Universidad de Salamanca
modesto@usal.es (ESPAÑA)

ELENA GIL

Universidad de Salamanca
elenagm@usal.es (ESPAÑA)

CRISTINA CALVO LÓPEZ

Universidad de Salamanca
cristinacalvolopez@usal.es (ESPAÑA)

Recibido: 15.12.2020

Aceptado: 16.12.2021

RESUMEN

Las redes sociales online se han ido convirtiendo en uno de los principales vehículos de comunicación y una de las mayores fuentes de información de actualidad. Esta creciente popularidad deja en evidencia la importancia de que los científicos sociales seamos capaces de analizar, interpretar y comprender en profundidad este nuevo tipo de herramientas. Este artículo tiene como objetivo mostrar los diversos métodos de análisis de la información pública obtenida a partir de una de estas redes, Twitter. Para ello tomamos como ejemplificación explicativa el caso #Cuéntalo, un episodio de narrativa compartida iniciado en esta red entre los días 26 y 28 de abril de 2018 tras la conocida sentencia de “La Manada”. A través de este caso se presentan aquí distintas metodologías para el estudio de los contenidos transmitidos, que van desde los análisis descriptivos más elementales hasta los análisis de contenido, pasando por la clasificación de

*Este trabajo ha sido posible gracias a la financiación recibida del Ministerio de Ciencia, Innovación y Universidades con fondos del Programa FEDER de la Unión Europea por el proyecto de referencia PGC2018-093755-B-I00. Asimismo, Cristina Calvo López agradece su ayuda para contratos predoctorales para la formación de doctores PRE2019-088733 financiada por MCIN (10.1303910/501100011033)

actores relevantes y el descubrimiento de la estructura de las relaciones entre los protagonistas y sus mensajes. Los resultados muestran cómo esta polémica sentencia derivó en una conversación digital viral donde distintas usuarias (en especial periodistas, escritoras y activistas feministas) comenzaron a compartir sus relatos de situaciones de violencia sexual vividas por las participantes o sus conocidas usando esta etiqueta, siendo capaces de identificar a las principales protagonistas, las distintas relaciones que establecieron entre ellas y sus mensajes y los principales temas que se conformaron en torno a ellos.

PALABRAS CLAVE

Redes sociales; Twitter; #Cuéntalo; análisis estructural; análisis dinámico, análisis de contenido.

ABSTRACT

Online social networks have become one of the main communication vehicles and one of the greatest sources of current information. This growing popularity shows the importance of social scientists being able to analyze, interpret and understand in depth this new type of tools. This article aims to show the diverse methods of analysis of public information obtained from one of these networks, Twitter. To do this, we take as an explanatory example the case of #Cuéntalo, an episode of shared narrative that began on this network between April 26 and 28, 2018 after the well-known sentence of “La Manada”. Through this case, we present different methodologies for the study of broadcasted content, ranging from the most elementary descriptive tools to content analysis, passing through the classification of relevant actors and the discovery of the structure of the relationships amongst their protagonists and their messages. The results show how this controversial sentence led to a viral digital conversation where different users (especially journalists, writers, feminists and influencers) began to share their stories of situations of sexual violence experienced by the participants or their acquaintances using this label. Through this analysis, it was possible to identify the main protagonists, the different relationships that they established between them and their messages and the main themes that were formed around them.

KEYWORDS

Social media, Twitter, #Cuéntalo, structural analysis, dynamic analysis, content analysis.

1. INTRODUCCIÓN

Es difícil imaginar en las sociedades modernas de hoy día un escenario en el que no medien las nuevas tecnologías. La comunicación se ha trasladado en muchos sentidos a un espacio virtual en el que la ciudadanía es continuamente emisora y receptora de mensajes multidireccionales, que además aparecen en contextos de continua mediación. Twitter es una plataforma que atrae la atención de millones de usuarios, incluidos los medios de comunicación que amplifican sus mensajes. Esta condición le otorga un papel relevante en muchas luchas sociales, configurándose como un nuevo espacio de información y socialización, cuyo análisis requiere la incorporación de nuevos métodos de análisis de datos masivos (*big data*). En este contexto, tiene más sentido que nunca incorporar a los estudios sociológicos los análisis de Internet, no solo para entender la movilización colectiva, sino para comprender más profundamente las dinámicas relacionales de la sociedad actual.

Mediante este artículo se pretende mostrar diversos modos de obtención y estudio de la información obtenida a través de Twitter, que van desde los análisis descriptivos más elementales hasta los análisis de contenido, pasando por la clasificación de actores relevantes y el descubrimiento de la estructura de las redes personales y temáticas. Para ello y a modo de ejemplificación explicativa, a lo largo de este artículo nos centraremos en el mecanismo clave para difundir la información: los *hashtags*, que permiten agrupar el contenido de las publicaciones que se comparten en la red social en torno a un tema concreto. Así, los mensajes son más visibles para los demás, de modo que estas etiquetas son muy útiles para la organización y promoción del activismo *online*. Adicionalmente, desde un punto de vista analítico, la selección de mensajes empleando una o varias etiquetas, le confiere un valor añadido por circunscribir la inmensidad de mensajes al estudio de uno o múltiples casos de interés. En concreto, este artículo se centra en el *hashtag* #Cuéntalo entre los días 26 y 28 de abril de 2018.

En esas fechas se generó una narrativa compartida en Twitter a raíz de la publicación de la polémica sentencia del caso “La Manada”, pues las redes se inundaron de mensajes de mujeres compartiendo sus experiencias como víctimas de violencia sexual. Mediante el análisis de estos mensajes se inducirá que muchas de las protagonistas denunciando del acoso fueron mujeres feministas denunciando muchos de los síntomas de la sociedad patriarcal vividos directa o indirectamente en carnes propias. Nadie puede negar que Internet refuerza y transmite los mensajes que circulan en los movimientos sociales, ofreciendo a los investigadores una nueva fuente de datos en las que construir evidencias empíricas para su conocimiento y análisis. Al hilo de este caso de estudio, constituido por la reacción a la sentencia de “La Manada” en la plataforma de Twitter, este artículo escrito para una revista de metodología contiene modos y técnicas de abordar la estructura y dinámica de los millones de mensajes que se difunden en la red a fin de potenciar que analistas de los fenómenos sociales empleen cada vez más este tipo de datos en sus investigaciones.

Resulta imposible negar que la popularización de Internet a amplias capas de la población posee una doble repercusión: por un lado, supone la decadencia de un modelo de construcción y divulgación de la opinión pública jerarquizado y controlado por unos centralizados medios de comunicación; por el otro, proporciona a los estudiosos de la sociedad de una nueva fuente para el estudio de la emisión y divulgación de las opiniones. Entre todas las plataformas y aplicaciones disponibles en la red de redes, la que más potencial contiene es Twitter tanto por los mensajes -de contenidos más públicos que privados- transmitidos por este medio, como por su estructura abierta que permite el acceso a su contenido incluso a aquellas personas que aún no se han registrado.

Además de la importancia para el estudio de la opinión pública, otro campo de estudio en el que Twitter presenta un gran potencial es en el de los movimientos sociales en la medida en que muchos de sus activistas y seguidores encuentran en sus tuits un poderoso medio tanto para difundir, recibir y criticar ideas como para movilizar a sus simpatizantes y divulgar sus logros.

El propósito de este artículo es doble en la medida en que trata describir cómo emerge pública y masivamente la indignación ante una sentencia judicial considerada a todas luces injusta mediante la denuncia pública de vejaciones que las mujeres han sufrido en su vida cotidiana, al mismo tiempo que pretende presentar sin pretensión de exhaustividad un elenco de posibles análisis que pueden llevarse a cabo con una extensiva muestra de mensajes con el fin de estimular la investigación sociológica sobre los mensajes que circulan en los nuevos medios sociales.

2. ANTECEDENTES

Ya son miles las publicaciones que se han realizado empleando Twitter como fuente de datos. Entre ellas prevalecen los estudios de opinión pública y comunicación, donde destacan los análisis sobre distintos procesos electorales y el modo de uso y difusión de los mensajes de candidatos y partidos (Larsson y Moe 2014, Casero-Ripollés 2020).

En esta línea predominan los análisis de contenido de los mensajes. No sucede lo mismo con los estudios sobre noticias falsas (Shao *et al.* 2018, Bovet y Makse 2019, Pérez-Dasilva *et al.* 2020), donde lo que suele interesar a los investigadores es localizar la procedencia y autoría de este tipo de informaciones para analizar su trayectoria viral, priorizando por ello la búsqueda de estructuras. También cabe destacar un tercer tipo de estudios, enfocados en los movimientos sociales (Freelon *et al.*, 2016, Jackson *et al.* 2018, Jost *et al.* 2018), donde se analizan tanto los vínculos entre los participantes como los contenidos de determinados fenómenos de protesta, como en el caso de Occupy Wall Street (Jensen y Bang 2013, Teocharis *et al.* 2015, Gleason 2013; Wang y Wang 2013). Asimismo, se encuentran investigaciones sobre las actitudes hacia sectores minoritarios como emigrantes y refugiados (Arcila *et al.* 2020, Gualda y Rebollo 2020)

Si bien son bastantes los estudios publicados sobre el #MeToo, precursor del #Cuéntalo, la gran mayoría de estas aproximaciones están realizadas desde enfoques tradicionales de la metodología de investigación social, como son las entrevistas en profundidad o las encuestas (Gill y Orgad 2018, Rodino-Colocino 2018, Kunst *et al.* 2019). Así, solo una pequeña parte de los análisis del #MeToo utilizan como base de datos este *hashtag* (Xiong *et al.* 2019, Deal *et al.* 2020, Jackson *et al.* 2020).

Ejemplos más similares al contenido de este artículo son el de Navarro y Coromina (2020) que estudian la reacción a la sentencia de la Manada con una perspectiva basada en el *framing* y la de Larrondo *et al.* (2019), quienes recogieron unos 291.000 tuits con la etiqueta #Cuéntalo y otras conectadas, con los que realizaron un análisis descriptivo seguido de un análisis hipotético-deductivo para determinar la relación entre el “hashtivismo”, el grado de politización y el comportamiento dialógico del feminismo en Twitter. Por su parte, Gutiérrez *et al.* (2020) estudiaron la reacción antifeminista a #Cuéntalo a través de un análisis estructural y un análisis de contenido, caracterizando los distintos contenidos que vertieron los hombres como respuesta a los mensajes de las experiencias relatadas por mujeres. También cabe mencionar otros estudios (Bucalo *et al.* 2020, Ruiz y Vallès 2020), que también usaron #Cuéntalo como base.

Este artículo tiene la particularidad de centrarse en una serie de propuestas técnicas y analíticas dirigidas a trabajar con grandes cantidades de datos desde una perspectiva analítica. Se examinan los métodos de extracción de tuits y se describe la información de Twitter disponible para su captura. Seguidamente, se aportan distintos análisis que pueden aplicarse a los tuits, clasificados en cuatro grandes tipos: descriptivo, dinámico, estructural y de contenido, usando la información extraída a partir de #Cuéntalo como ejemplo de cada uno.

3. LA EXTRACCIÓN DE LOS DATOS.

Para la extracción de datos de Twitter existen diversas opciones. La API de esta plataforma (<https://developer.twitter.com>) permite tanto recoger mensajes ya publicados como prepararse para descargarlos en directo (*streaming*). Del primer modo se hace una petición (*query*) según un criterio específico para descargar los tuits escritos hasta la fecha que tengan una palabra o etiqueta (*hashtag*) acompañada o no de criterios complementarios. En el modo en directo se pueden obtener todos aquellos tuits que se escriban a partir del momento del inicio de la recolección siguiendo los criterios deseados. Dado que se generan muchos tuits a lo largo del día, los criterios que se utilicen para efectuar una extracción de un tuit grabado o en directo dependerá de los objetivos perseguidos. Sea por las limitaciones de la API pública, sea por los emisores o por los temas, siempre habrá límites temporales: el comienzo de la recolección y el final de ella.

Un criterio útil es el relacionado con los *usuarios*: puede ser interesante ver cuál es el comportamiento comunicativo de instituciones o personas. También puede ser un criterio la temática. Para ello se puede recurrir a frases o *palabras*

(con sus sinónimos), y especialmente a las *etiquetas*. Este último sistema es probablemente el más empleado en la investigación académica para circunscribir las muestras. Es posible unir dos criterios, el tiempo y el tema, como conviene al objetivo de este artículo: analizar el uso de la etiqueta #Cuéntalo durante sus días de apogeo, entre el 26 y 28 de abril de 2018.

Existen numerosos programas para realizar estas descargas. Destacan las herramientas complementarias contenidas en los CAQDAS (NVivo, Atlas-ti, QDA), el paquete rtweet en R (Kearney 2019), el T-Hoarder de Congosto *et al.* (2017) en Linux y la herramienta web DiscoverText (Shulman 2011). Para esta investigación hemos utilizado las dos últimas herramientas. Buscando el *hashtag* #Cuéntalo en las fechas indicadas, obtuvimos dos bases de datos semejantes de 50.000 tuits y unos 450.000 retuits cada una. Para tratarlos se ha elaborado un paquete propio en R al que se ha denominado *tweetCoin*, que está basado en el más extendido *rtweet* (Kearney 2019).

Antes de proseguir cabe señalar que la selección de tweets no es un proceso aleatorio. Está circunscrito a los usuarios, las etiquetas, los textos o los *hashtags*. Conviene, pues, precisar que las investigaciones de este tipo están acotadas en un tiempo, tema, lengua y/o emisores.

4. LA INFORMACIÓN DE LOS TUIITS.

Aunque los nombres pueden variar dependiendo de la aplicación que se emplee, la información que contiene cada tuit puede clasificarse en los siguientes tipos: a) información de los mensajes, b) atributos del autor y c) variables derivadas de los registros de tuits.

Tabla 1. Información de los tuits: relacionadas con los mensajes.

Atributo	Descripción
<i>tweet_id</i>	Es un número único para cada tuit, consta de más de dieciocho dígitos.
<i>text</i>	Registro central. Contiene los 140 o 280 caracteres de mensaje de Twitter.
<i>retuits</i>	El número de retuits (<i>retweets</i>) recibidos hasta el momento.
<i>favoritos</i>	El número de favoritos (o “me gusta”) que ha recibido.
<i>author</i>	Contiene el nombre del usuario que ha escrito el mensaje, si es original, o del que lo ha reenviado, en el caso de que se trate un retuit.
<i>date</i>	La fecha en la que se ha emitido el mensaje.
<i>geolocation</i>	La geolocalización hace alusión al lugar desde que se ha enviado un mensaje. Solo una ínfima minoría la contienen.
<i>source</i>	Es el dispositivo desde el que se manda. Casi todos los mensajes contienen esta información.
<i>type_media</i>	Campo que indica que tipo de contenidos adicional aparece en un tuit.
<i>url_media</i>	Lugar donde reside su contenido adicional.
<i>lang</i>	El idioma en que está escrito el mensaje.
<i>link</i>	Lugar donde puede localizarse el mensaje.

FUENTE: Elaboración propia a partir de los nombres de los vectores de los objetos del paquete rtweet.

Entre la información derivada de los tuits (tabla 1) cabe destacar *text*, *source* y *date*. En ocasiones también puede ser interesante emplear información sobre la geolocalización, desde qué dispositivo se ha enviado, el idioma en el que está escrito o el tipo de medio que lo acompaña (foto, vídeo, enlace, ...).

Tabla 2. Información de los tuits: atributos del autor.

Atributo	Descripción
<i>user_id</i>	Identidad del usuario. Número que lo identifica permanentemente.
<i>user_name</i>	Identificativo con el que se le puede citar un perfil (“@”).
<i>screen_name</i>	Nombre completo del usuario de identificación.
<i>location</i>	Lugar al que se adscribe un usuario. No tiene que ser una localidad real.
<i>description</i>	Párrafo en el que el usuario escribe sobre su identidad.
<i>url</i>	Dirección web a la que el usuario redirige (otra red social, página personal...). Su inclusión es
<i>followers</i>	Número de cuentas que siguen al emisor del mensaje.
<i>following</i>	Número de personas que sigue en Twitter el emisor.
<i>listed</i>	Número de listas del emisor.
<i>created_at</i>	Fecha en la que el usuario se dio de alta en Twitter.
<i>favorite_count</i>	Número de usuarios favoritos que dispone cada usuario.
<i>profile_image_url</i>	Dirección que contiene la foto del perfil del usuario.

FUENTE: Elaboración propia a partir de los nombres de los vectores de los objetos del paquete rtweet.

De la información del autor de cada tuit (tabla 2) destacan el *user_name* que es el usuario, por ejemplo @Lafallaras, y que coincide con lo que en mensajes (tabla 1) se denominó *author*. El *screen_name*, por otro lado, hace referencia al nombre público del usuario, como Cristina Fallarás. También es interesante *description*, un texto en el que el usuario describe su identidad (real o no). De los perfiles se extraen también otros componentes, como el número de *followers* (seguidores) y de *following* (seguidos). El interés del resto de componentes dependerá del objeto de la investigación.

Tabla 3. Información de los tuits: derivadas del registro de la información.

Tipo	Atributo	Descripción
a) Usuarios	<i>Usuarios mencionados</i>	Todas las menciones (expresadas con una “@” delante del <i>screen_name</i>) de aquellas personas que se quiere aludir en el mensaje original
	<i>Usuarios retuiteados</i>	En todo mensaje retuiteado se coloca delante el acrónimo RT seguido por el <i>screen_name</i> de quien lo emitió originalmente.
b) Temática	<i>Primer hashtag</i>	Las etiquetas (“#”), son palabras o composiciones que permiten agrupar y destacar el contenido de un mensaje.
	<i>Total hashtags</i>	Puede tenerse en cuenta el número de <i>hashtags</i> , sólo el primero (más importante) o todos los que aparezcan.
c) Mensajes	<i>Contenido multimedia</i>	Pueden extraerse los contenidos multimedia (fotos, videos, páginas web, ...) que los usuarios incluyen en los mensajes.
d) Popularidad	<i>Tipo de usuario</i>	Hay diferentes clasificaciones de usuarios según su popularidad. Destaca la de González-Bailón (2017).

FUENTE: Elaboración propia a partir de los nombres de los vectores de los objetos del paquete rtweet.

Finalmente, entre la información derivada del registro de la información de los tuits, destacan cuatro tipos de datos: los relacionados con usuarios, los vinculados con la temática, los que contienen los mensajes y los asociados a la popularidad del emisor (tabla 3).

5. ANÁLISIS DESCRIPTIVO: QUIÉN ESCRIBE Y QUÉ TEMAS RESALTA.

Una vez descargados un conjunto de tweets con una determinada etiqueta (o cualquier otra condición), habría que separar aquellos que son originales de los que son reenvíos. En el caso que nos ocupa, se obtuvieron 52.411 mensajes originales con la etiqueta #Cuéntalo. Pero al mismo tiempo es importante añadir los 424.949 reenvíos, las 2.792 citas y las 11.106 respuestas, es decir, un total de 491.258 intervenciones procedentes de un total de 162.172 perfiles distintos en un periodo limitado a tres días, es decir, estamos ante un fenómeno digital de relevancia.

En función de la información disponible en los mensajes de Twitter, es conveniente comenzar con unos descriptivos preliminares de sus elementos centrales, distintos del texto, es decir, *autor* del mensaje original, el perfil que realiza la *reemisión* del mensaje los *hashtags* y las *menciones* contenidos ambos en el texto con los prefijos respectivos de # y @, para lo que han de emplearse funciones que los aislen.

Observando la frecuencia de los mensajes originales emitidos pueden descubrirse los perfiles más activos, entre los que se encuentra la periodista del Diario Público Cristina Fallarás (@LaFallaras), iniciadora de los mensajes. Le siguen dos cuentas: @MujeresUC3M, perteneciente a la Asociación de Mujeres de la Universidad Carlos III de Madrid y el denominado @SoldadoDescono3, cuenta actualmente suspendida, que actuó como reacción machista contraria a #Cuéntalo.

También es importante indagar qué perfiles se mencionan por quienes escriben los mensajes. Aunque la gran mayoría de los mensajes no mencionan a nadie, queda claro de nuevo que la clara protagonista fue @LaFallaras, no solo por ser la iniciadora de la etiqueta y quien más mensajes originales emitió, sino también por ser con diferencia la usuaria más invocada por otros perfiles. Otros perfiles mencionados fueron los de Barbijaputa (@barbijaputa), La Vecina Rubia (@lavecinarubia), Beatriz Talegón (@BeatrizTalegon), Irene Montero (@irene_montero_) y Virginia P. Alonso (@Virginiapalonso), todos correspondientes a periodistas, activistas y políticas de gran popularidad en la red, que ayudaron a promover la expansión de los mensajes. Asimismo, con 44 y 39 menciones respectivamente aparecen los dos periódicos más implicados en la denuncia: @eldiarioes y @publico_es, en los que escribían @barbijaputa en el primero y @LaFallaras y @Virginiapaolsonso en el segundo.

Tabla 4. Frecuencia de mensajes originales emitidos por los usuarios más activos.

Perfiles	Emitidos	Mencionados
@LaFallaras	242 Tuits sin menciones	36804
@MujeresUC3M	133 @LaFallaras	1170
@SoldadoDescon3	129 @Barbijaputa	156
@Hesperion	84 @lavecinarubia	106
@ttesbot	84 @BeatrizTalegon	63
@dramalesbian	80 @irene_montero	61
@trendinaliaES	77 @Virginiapalonso	57
@calvall00	72 @eldiarioes	44
@EstherGRecuero	65 @policia	44
@AnuncioPublico1	61 @srtabebi	42
@fuetendencia	61 @publico_es	39

FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017).

En otro orden de elementos analizables, el uso de *hashtags* refleja el interés por parte de los usuarios de que sus mensajes se relacionen con la etiqueta en cuestión (Pérez-Martínez y Rodríguez-González, 2017). Además, como en un mismo mensaje se puede expresar más de uno, su análisis permite ver cómo los temas de interés están interconectados. De esta forma, el interés de las etiquetas radica en conocer las co-apariciones, averiguar cómo se relacionan entre sí los distintos *hashtags* y detectar conjuntos de temas. Esto es de especial importancia cuando se trabaja con una base de datos compuesta por diversos *hashtags* no sinónimos, de manera que pudiéramos entender los denominadores comunes de cada tema.

En este conjunto seleccionado por su etiqueta, el *hashtag* más frecuente es, como cabe esperar, #Cuéntalo. En la tabla 5 aparecen ordenados por frecuencia otros *hashtags* asociados, entre los que destacan #noesno, #yositecreo, #lamana y #noesabusoesviolacion, etiquetas relacionadas con la indignación por la sentencia de La Manada (Larrondo *et al.* 2019, Gutiérrez *et al.* 2020). Todo indica que hay otras que no pertenecen al tema: #graciasiniesta y #wwegrr proceden de robots de tendencias como @ttesbot (línea 5 de la tabla 4) que recopilan las etiquetas más usadas en cada momento.

Tabla 5. Frecuencia de las etiquetas que acompañaron a #Cuéntalo (originales).

	Frecuencia
#cuentalo	52411
#noesno	1319
#yositecreo	857
#lamanada	821
#felizsabado	311
#noesabusoesviolacion	304
#metoo	214
#graciasiniesta	166
#wwegr	162
#justiciapatriarcal	142

FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017).

Volviendo a los perfiles, los más destacados hasta el momento podrían clasificarse en cinco clases: personales, grupales, mediáticos, robots y suspendidos. Ahora bien, más interesante resulta clasificarlos en función de dos dimensiones: *actividad* y *popularidad*, que pueden ser medidas tanto en el contexto analizado, como en el conjunto general de la participación en Twitter. Indicadores de actividad son las *emisiones* (suma de tuits y retuits) en un determinado contexto y, en general, el *estatus* (número de mensajes retransmitidos desde el inicio de una cuenta) y el número de perfiles que se siguen (*following*). Indicadores de popularidad son el número de *retuits recibidos* en el contexto y el número de personas que siguen a un determinado perfil (*followers*). En la tabla 6 aparecen todos estos indicadores ordenados por el número de emisiones. Sobresale de nuevo @LaFallaras, seguida de @dramalesbian_ con un equilibrio de mensajes originales y compartidos de otras cuentas (retuits enviados).

Popularidad y actividad pueden combinarse para generar una tipología de usuarios (González-Bailón *et al.* 2013). En el gráfico 1 puede observarse cómo la primera se distribuye horizontalmente en función del número de seguidores frente a los seguidos, y la segunda verticalmente según los mensajes emitidos frente a los mensajes divulgados, generando cuatro tipos.

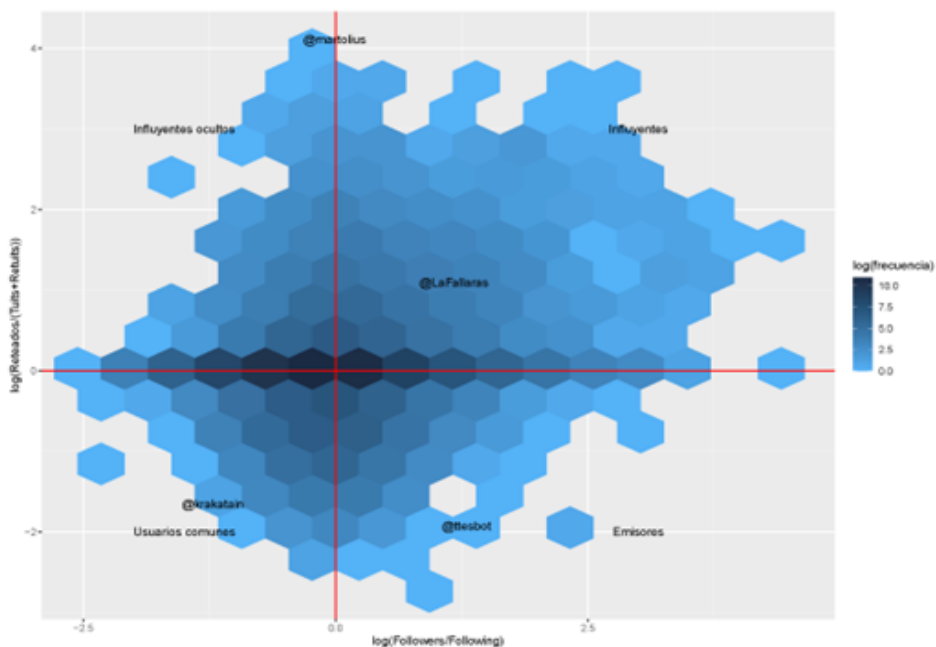
Tabla 6. Estadísticas de los perfiles más activos y/o mencionados en el contexto de #Cuéntalo

Perfil	Clase	Contexto				General			
		Actividad		Popularidad		Actividad		Popularidad	
		Emisiones	Tuits	Retuits enviados	Retuits recibidos	Estatus	Sigue a	Seguido por	Tipo
@LaFallaras	Personal	546	242	304	6.785	26.988	4.377	65.018	Influyente
@dramalesbian_	Personal	168	86	82	451	583.061	513	10.085	Influyente
@SoldadoDescono3	Suspendida	146	129	17	230	41.022	2.063	2.081	Influyente
@MujeresUC3M	Grupal	144	133	11	313	1.854	249	1.449	Influyente
@ttesbot	Robot	84	84	0	1	250.106	824	16.298	Emisor
@caval100	Personal	79	72	7	106	503.202	95.027	125.070	Influyente
@trendinaliaES	Robot	77	77	0	11	795.608	48	49.813	Emisor
@AnuncioPublico1	Robot	61	61	0	0	60.804	20	27.672	Emisor
@fuetendencia	Robot	61	61	0	1	58.341	2	531	Emisor
@Virginiapalonso	Personal	23	5	18	502	43.929	1.973	30.983	Influyente
@Barbijaputa	Personal	13	8	5	2.154	127.185	944	300.540	Influyente
@publico_es	Prensa	13	8	5	848	294.957	1.418	897.748	Influyente
@sratabebi	Personal	9	7	2	1.390	32.405	335	530.983	Influyente
@BeatrizTalegon	Personal	3	3	0	648	66.519	18.340	172.636	Influyente
@lavencinarubia	Personal	3	3	0	5.010	45.078	911	424.853	Influyente
@eldiarios	Prensa	2	2	0	250	150.071	463	891.948	Influyente
@Irene_Montero_	Personal	2	1	1	6.992	9.599	1.804	237.157	Influyente

FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017).

De estos, dos son activos: los *emisores*, que emiten más mensajes de los que les divulgan, y los *influyentes*, en los que confluyen actividad y popularidad. Otros dos son pasivos: los *influyentes ocultos*, populares e inactivos, frente a los *usuarios comunes*, que no son ni activos ni populares. Cabe destacar que todos los perfiles listados en la tabla 6 pertenecen a los dos primeros tipos, puesto que solo recogen aquellos que han sido muy activos en el contexto de #Cuéntalo.

Revisando otros perfiles no citados hasta este momento, caben destacar los grandes periódicos nacionales, ya que, con la excepción de ABC, todos participaron con la emisión de tuits. También debe mencionarse los perfiles de Rosa María Palacios (@rmpalacios), periodista peruana que interviene animando a visitar el *hashtag*; el de Íñigo Errejón (@ierrejon) expresando “espanto” y “rabia” al entrar en #Cuéntalo, y el de Juan Carlos Monedero (@monederoJC) reflexionando sobre el poco respeto de los hombres hacia las mujeres e invitando a leer los mensajes de @LaFallaras. Además de estos, aparecerán otros distintos en el análisis estructural por haber sido objetos de numerosos reenvíos de sus mensajes originales.

Gráfico 1. Distribución de los índices de emisión (vertical) y seguimiento (horizontal)

FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto et al. 2017).

6. ANÁLISIS DINÁMICO.

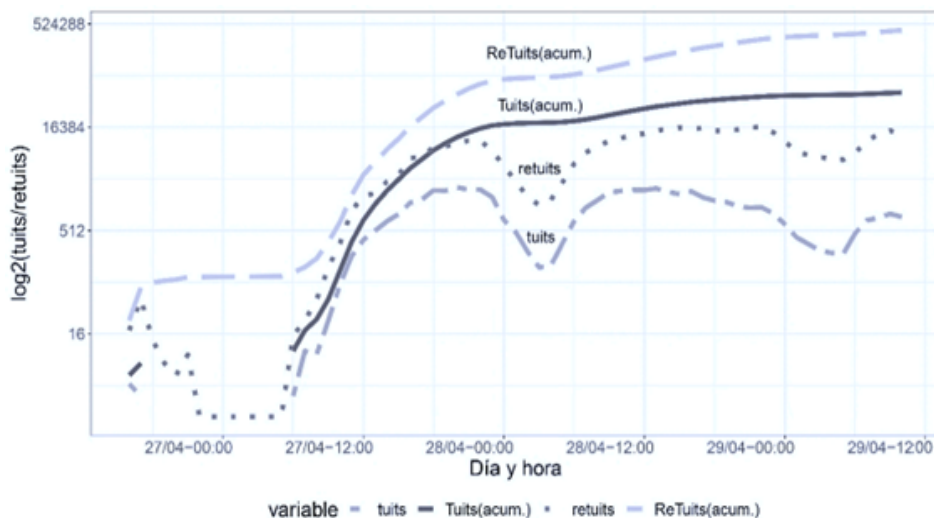
Twitter ofrece un diseño idóneo para el estudio regular y continuado de la evolución de un segmento importante de la opinión pública. Contar con todas y cada una de las unidades emitidas a través de este canal confiere una gran ventaja para estudiar la difusión de las opiniones tanto a corto, como se hace en el ejemplo de este artículo, como a largo plazo. En este sentido, el análisis dinámico implica los mismos procedimientos que el análisis estático, pero añadiendo el componente temporal.

6.1. Frecuencia temporal de los mensajes.

Los tuits recogidos contienen la fecha de emisión expresada en día y hora. Por ello, para ver cómo se distribuyen los mensajes a lo largo del tiempo hay que recodificar esta variable en una unidad de tiempo mayor que el segundo. Dependiendo del alcance temporal del estudio, esta unidad puede ser la hora, el día, la

semana o, para estudios más extensos, varios meses o el año. En el gráfico 2 se representan las frecuencias simples y acumuladas de tuits y retuits en una escala logarítmica, donde cada línea vertical representa 6 horas de tiempo.

Gráfico 2. Número de tuits y retuits por hora (simples y acumulados)



FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017).

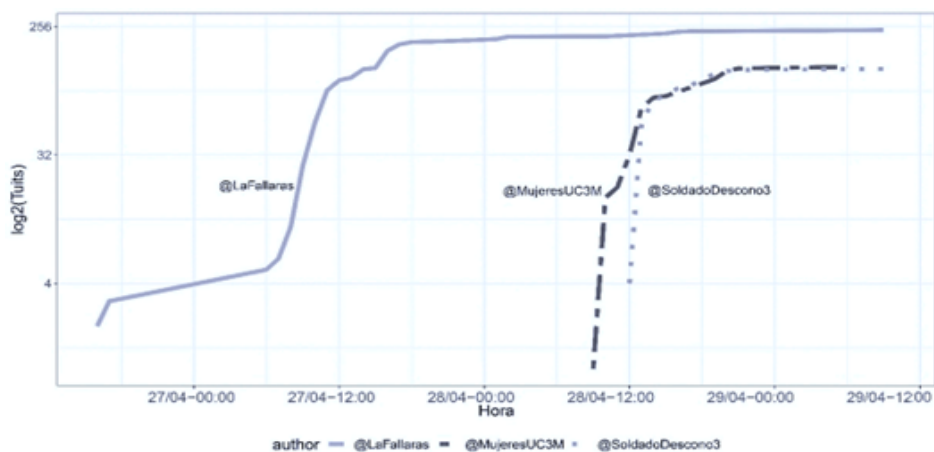
Aunque los primeros tuits con el *hashtag* #Cuéntalo comenzaran el día 26, llegándose a contabilizar un centenar de retuits, el gráfico muestra de modo claro que el fenómeno comienza a adquirir el carácter viral a partir de las seis de la mañana del día 27, alcanzando su pico diario en torno a las 10 de la noche. Vuelve a repuntar en torno a las cuatro de la mañana del día 28, y en la madrugada del 29 se nota un descenso menor que en la noche anterior, lo que es posible señal de la incorporación de perfiles del mundo hispanoamericano, donde también el tema se hizo viral algo más tarde.

Es de interés conectar esta evolución con las manifestaciones masivas que tuvieron lugar en la tarde del día 26 en contra de la sentencia. Aunque el inicio de los mensajes con esta etiqueta fue anterior, la masificación de los mensajes solo tuvo lugar a partir del día siguiente, por lo que se podría conjeturar que la manifestación fue catalizadora de la eclosión de los mensajes en Twitter más que en el sentido contrario.

6.2. EVOLUCIÓN TEMPORAL DE LOS AUTORES

Resultaría complejo representar en un solo gráfico la evolución temporal de todos y cada uno de los usuarios que emiten el conjunto de los mensajes involucrados en un contexto. Por ello, solo se representan en el gráfico 3 los tres perfiles que emitieron más de un centenar de tuits originales. Estos son @LaFallaras, @MujeresUC3M y @SoldadoDescono3.

Gráfico 3. Número de tuits acumulados lanzados por los perfiles @LaFallaras, @MujeresUC3M y @SoldadoDescono3



FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto et al. 2017).

1) La líder con diferencia en envío de mensajes originales es @LaFallaras, con un total de 242. En menos de una hora en la tarde del jueves 26 lanza tres mensajes, donde comparte los testimonios de experiencias de acoso y agresiones sexuales de otras tres mujeres acompañándolas del *hashtag*.

2) Otra cuenta prolífica en el envío de mensajes fue @MujeresUC3M. Se trataba de una cuenta de una asociación de mujeres universitarias, actualmente @AUCattana, que sirvió de altavoz para que aquellas mujeres que por una razón u otra no quisieran compartir públicamente sus vivencias, tuvieran un medio para hacerlo.

3) El tercer perfil, con más de cien mensajes originales, fue @SoldadoDescono3. Es una cuenta actualmente suspendida de claro contenido misógino y antifeminista cuyo primer mensaje remite a una página antifeminista llamada “A Voice for Men” (<https://avoiceformen.com>) que dice convocar a un “Movimiento a favor de los Derechos de los Hombres”. El resto de sus mensajes usando #Cuéntalo son una retahíla de comentarios satíricos con la intención de ridiculizar el movimiento y a sus participantes a través de mensajes narrando si-

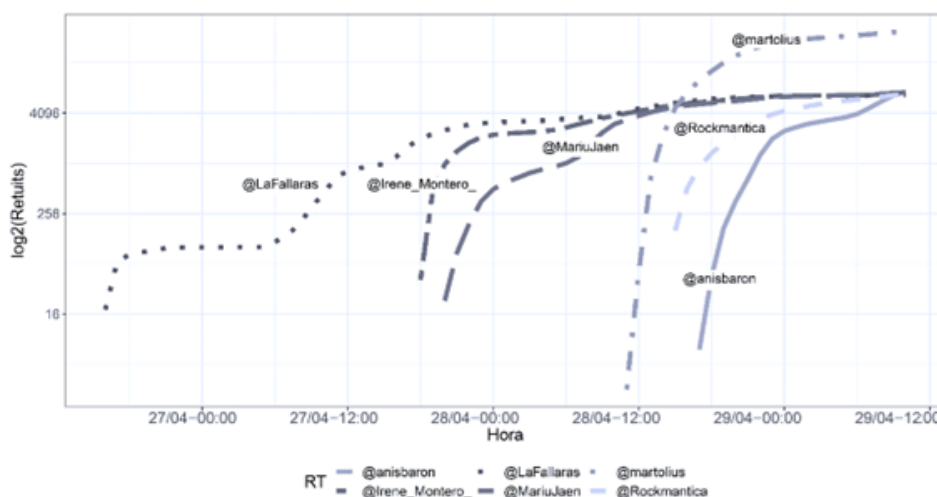
tuaciones absurdas, queriendo mostrar irónicamente el contenido de los mensajes de las mujeres.

Lo curioso de este perfil es que aparece un par de horas después de los mensajes de las mujeres de la Carlos III, e iguala el número de tuits de esta, aunque se queda lejos en el número de retuits recibidos. En ambos casos, su actividad se concreta el sábado y en las primeras horas del domingo prácticamente desaparecen, como muestra el gráfico 3.

6.3. Evolución temporal de los retuits recibidos

Al analizar la evolución del número de retuits recibidos aparece por primera vez uno los perfiles más relevantes de #Cuéntalo: @martolius, una usuaria con un número no muy alto de seguidores. La diferencia entre el número de retuiteos que recibe @martolius respecto a las cinco cuentas con más mensajes compartidos es más que recalcable, como puede comprobarse en el gráfico 5 acumulado.

Gráfico 5. Número de retuits acumulados por hora



FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto et al. 2017).

Lo interesante de este análisis dinámico consiste en observar cuándo empiezan a hacerse populares los mensajes y hasta qué punto esta relevancia se mantiene. Como era de esperar, la trayectoria más continuada es de @LaFallaras. Aunque hasta la mañana del viernes sus mensajes no suben popularidad, recobra viralidad de modo progresivo hasta lograr estabilizarla. @Irene_Montero_ y @MariuJaen aparecen el viernes por la tarde, consiguiendo mantenerse hasta el final de la serie. Por su parte, @martolius aparece el sábado antes del mediodía,

y @Rockmantica y @anisbaron llegan las últimas, a pesar de que sus mensajes llegan a superar la popularidad de @Irene_Montero_ y @MariuJaen.

Como se ha podido comprobar, el análisis temporal puede hacer algo muy potente: detectar cuándo y a través de quiénes se viraliza un *hashtag*. En este caso, fue lanzado por una periodista y potenciado tanto por feministas influyentes en tres vertientes: periodística, política y tuitera en general, como también por perfiles de la esfera política (Podemos) y de los medios de comunicación (Público y elDiario.es, sobre todo). Finalmente, fue replicado y fomentado por usuarias comunes.

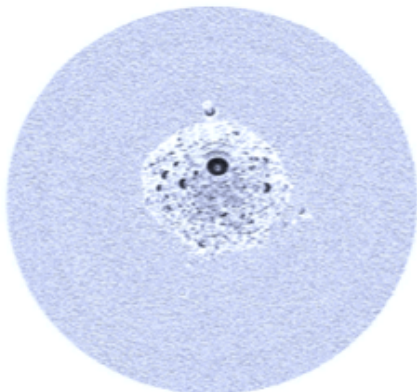
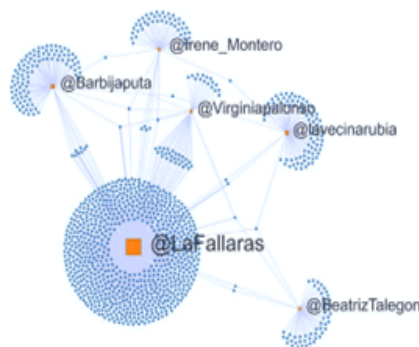
7. ANÁLISIS ESTRUCTURAL

Un análisis estructural es aquel que estudia los sistemas de relaciones entre elementos. A la hora de enfrentarse a los elementos de los mensajes de Twitter –sin prejuicio de otros–, resaltan por su importancia tres tipos de análisis estructurales, a los que podría añadirse el constituido con las palabras contenidas en los mensajes. El primero de ellos son las menciones, el segundo los retuits y el tercero los *hashtags*.

En general, los elementos de una estructura pueden ser representados mediante un grafo. En este, los nodos representan cada uno de los componentes de esa estructura y los enlaces las relaciones (menciones, retuits, coocurrencias). La disposición de los primeros se realiza conforme a un algoritmo, entre los que son más frecuentes el Fruchterman-Reingold para redes pequeñas, y el ForceAtlas-2 eficiente en el caso de redes con un elevado número de elementos. Estos procedimientos son de tal naturaleza que ubican los nodos más relacionados en el centro, aglutinándose entre ellos de modo que conformen grupos de relaciones mutuamente exclusivas, llamadas comunidades, que se obtienen mediante complejos procedimientos matemáticos.

7.1. Estructura de las menciones.

En el gráfico 6 se muestra el conjunto de menciones en tuits originales. En él destaca la figura central de @LaFallaras. Apenas destacan también otros perfiles mencionados debido a las limitaciones de visualización del papel. Para ello, la red ha de simplificarse: un procedimiento posible es eliminar aquellos perfiles que no se conectan con las seis grandes cuentas mencionadas como puede verse en el gráfico 7. Todas ellas están intermediadas por otras usuarias, especialmente @Virginiapalonso y @LaFallaras. Al haber sido elaborados con las menciones, los gráficos 6 y 7 dan cuenta del liderazgo de los emisores: mientras mayor número de menciones reciban, mayor centralidad adquirirán en la representación.

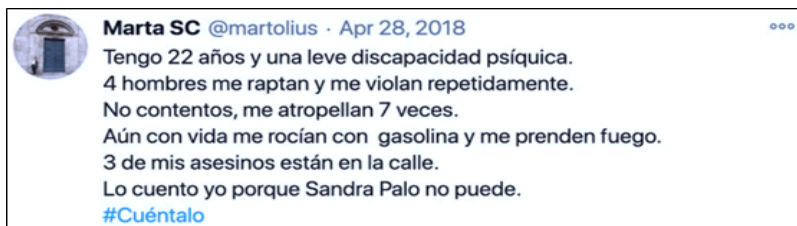
Gráfico 6. Red de menciones con la etiqueta #Cuéntalo**Gráfico 7. Red de menciones centrales con la etiqueta #Cuéntalo**

FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto et al. 2017).

7.2. Estructura de los retuits.

Otra perspectiva distinta de las menciones la proporcionan los retuits. Mientras que en una mención el perfil activo escribe un mensaje original e incluye en el texto el perfil de aquellas personas que considera implicadas en su contenido, en un retuit el perfil activo hace uso del mensaje de otra persona y lo incluye en su propio perfil con el objeto de difundirlo. De esta forma, realizando una red de retuits se puede observar la importancia que se ha otorgado a cada mensaje.

Si nos fijamos en las frecuencias de perfiles retuiteados (centralidad de grado de entrada en la terminología de grafos), se observa en la tabla 7 la gran atracción del mensaje de @martolius. En menos de una hora, el sábado por la mañana, escribió dos mensajes usando #Cuéntalo, siendo el segundo retuiteado en 24 horas por casi 40.000 perfiles distintos. Su texto era el siguiente:

Imagen 1 Mensaje de @martolius.

FUENTE: Twitter.

También destacan, con más de 7.000 retuits recibidos, los mensajes de @anisbaron y los de @MariuJaen. En la tabla 7 también se encuentra @LaFallaras, quien redactó 242 mensajes distintos y emitió 304 retuits (grado de salida). Consecuencia de ello, se convirtió con diferencia en la gran intermediadora popular de la red. Su bajo autovalor revela que fueron otros perfiles (@martolius y @anisbaron) los que estuvieron más relacionados con los nodos centrales.

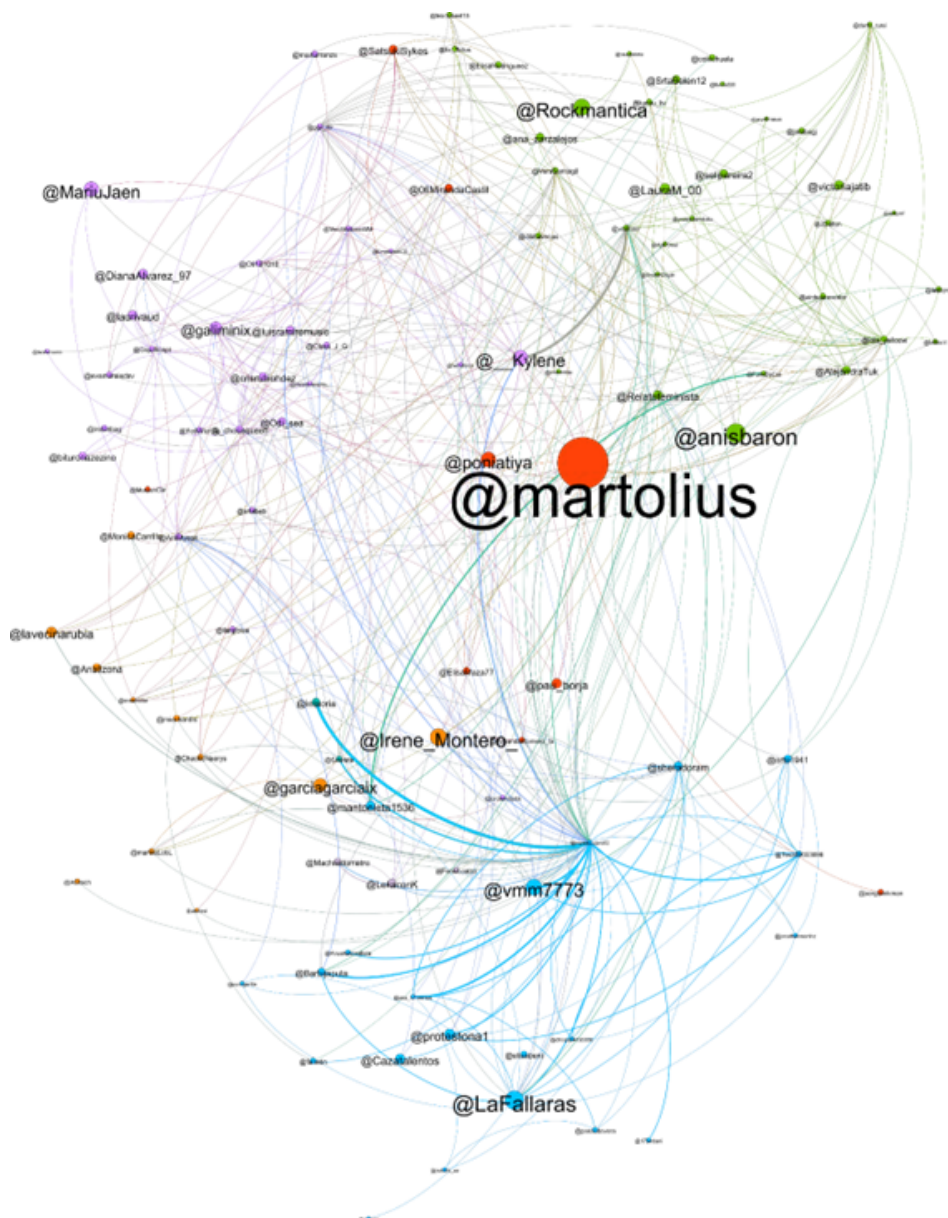
Independientemente del número de tuits disponibles, resulta interesante realizar una representación de la estructura de las relaciones principales seleccionando solo aquellas con un grado de entrada mayor que un determinado valor¹. Además de ello, al conjunto del grafo se le aplica previamente el cálculo de sus comunidades mediante la modularidad de sus nodos (Jacomy *et al.* 2014). Así, puede obtenerse un mapa de los distintos sectores participantes en un contexto (gráfico 8).

Tabla 7. Indicadores de centralidad por perfil

	Grado entrada	Grado salida	Inter- mediación	Autovalor
@martolius	38.890	1	0,00	1,00
@anisbaron	7.392	7	0,11	0,12
@MariuJaen	7.261	3	0,10	0,07
@Irene_Montero_	6.992	1	0,12	0,07
@Rockmantica	6.847	0	0,00	0,09
@LaFallaras	6.785	304	1,00	0,06
@garciagarciaix	6.224	3	0,08	0,05
@poniatiya	5.998	4	0,08	0,09
@vmm7773	5.353	2	0,10	0,06
@lavecinarubia	5.010	0	0,00	0,03

FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017). Estos indicadores han sido obtenidos con el paquete igraph de R. La intermediación se ha normalizado con límites 0 y 1.

¹ El gráfico completo de retuits puede verse en Congosto (2018).

Gráfico 8. Estructura de mensajes republicados. (Selección de nodos por grado)

FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017).

7.3. Estructura de los hashtags.

Adentrándose en el próximo apartado sobre el contenido de los mensajes, el tercer elemento sobre el que se pueden construir estructuras son los *hashtags*. El interés aquí radica en conocer cuáles son las etiquetas que acompañan a las publicaciones en el conjunto de la muestra. Como explicamos antes, todo el conjunto que se está analizando contiene la etiqueta #Cuéntalo, pero es habitual que los mensajes incluyan más de una. En la ya analizada tabla 5 aparecía la etiqueta #Cuéntalo en la totalidad del conjunto, junto con el resto de las etiquetas con menor frecuencia. Pero aquella tabla carecía de elementos relacionales, para lo cual habría que cruzar la ocurrencia de las etiquetas. En la tabla 8 aparecen las frecuencias cruzadas de #Cuéntalo con las otras diez etiquetas más frecuentes.

Ahora bien, si se desea una visión conjunta de toda la estructura, más que una tabla repleta de categorías resulta más conveniente representar toda esta información en un grafo. En este caso, los nodos serían cada una de las etiquetas que estarían unidas, siempre y cuando el residuo normalizado fuera positivo y significativo —por encima de 1.64 en una distribución normal si se adopta el criterio de una hipótesis unidireccional con un nivel de significación del 5%— (Escobar y Martínez-Urbe 2020). Partiendo de la tabla 8 extendida a las 278 etiquetas que han aparecido más de una vez en los mensajes originales recopilados, se obtuvo la red expuesta en el gráfico 9.

Tabla 8. Frecuencias de co-aparición de etiquetas en los mensajes con #cuéntalo

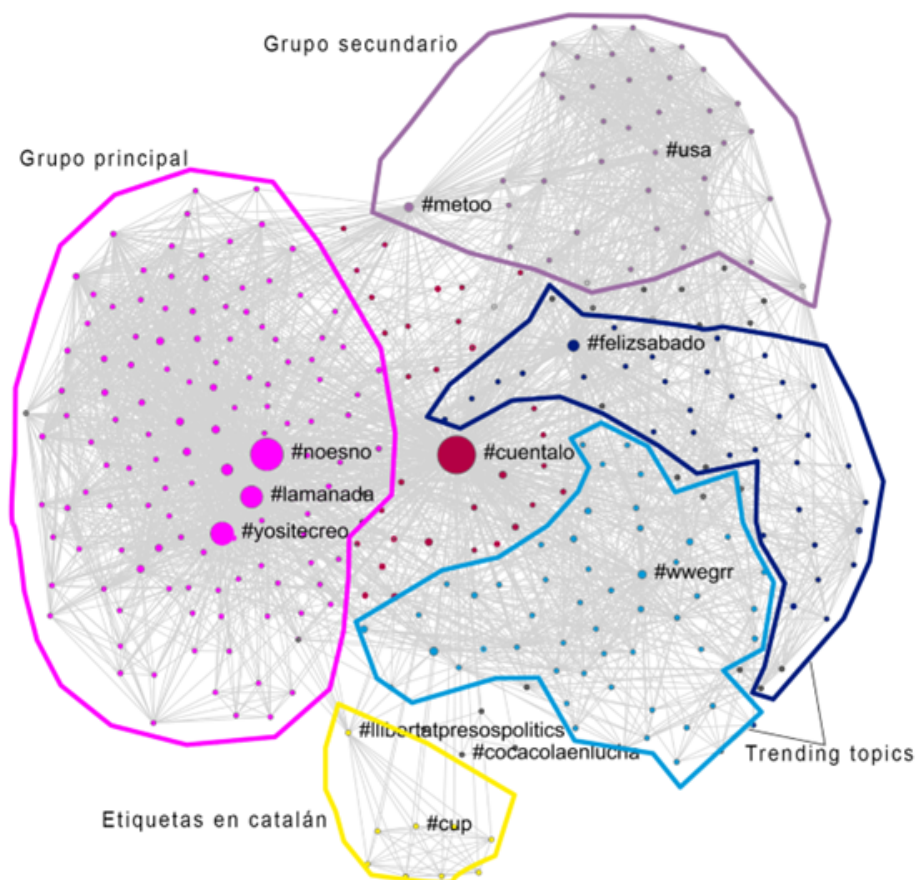
		A	B	C	D	E	F	G	H	I	J
#cuentalo	A	52411	1317	855	818	307	303	214	166	158	142
#noesno	B	1317	1319	322	250	12	191	19	4	1	53
#yositecreo	C	855	322	857	220	10	107	18	4	1	68
#lamanada	D	818	250	220	821	18	101	12	1	1	63
#felizabado	E	307	12	10	18	311	2	8	0	0	3
#noesabusoesviolacion	F	303	191	107	101	2	304	11	2	1	41
#metoo	G	214	19	18	12	8	11	214	0	0	3
#graciasiniesta	H	166	4	4	1	0	2	0	166	66	2
#wwegrr	I	158	1	1	1	0	1	0	66	162	1
#justiciapatriarcal	J	142	53	68	63	3	41	3	2	1	142

FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017).

En ese grafo se pueden distinguir distintos grupos de nodos o etiquetas formado por uno de los múltiples procedimientos existentes para conformar comunidades en las redes. En este caso, fue elegido el denominado Lovaine (Blondel 2008). En él podemos distinguir cinco comunidades claras: un grupo temático principal representado por los hashtags #noesno, #lamanada y #yositecreo, uno secundario donde destacan #metoo y #usa, dos grupos formados a partir de los

trending topics y un último en el que queda representada la comunidad tuitera catalana a través de etiquetas como #cup o #llibertatpresopolitics.

Gráfico 9. Gráfico de comunidades



FUENTE: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017)

Un elemento común de las dos últimas redes presentadas es el uso de comunidades, que poseen la funcionalidad de distinguir grupos en un conjunto de relaciones. Existen múltiples procedimientos para su descubrimiento, pero exponerlos en este artículo queda fuera del alcance de su extensión. Una buena introducción es el trabajo de Porter *et al.* (2009). La investigadora que desee (re) construir y mostrar análisis estructurales puede recurrir a muy diversos programas para llevarlos a cabo. Aunque sean de pago, los más empleados son Ucinet y NodeXL, también se puede recurrir a alternativas gratuitas como son el caso de Pajek, Gephi y los paquetes de R o Python como son igraph y network (Net-

workX). Para esta investigación se han empleado paquetes propios como el ya disponible netCoin y otro en desarrollo que se denomina tweetCoin.

8 ANÁLISIS DE CONTENIDO

Desde que Thomas y Znaniecki emprendieron en 1918 el análisis de la integración de los emigrantes polacos en Estados Unidos mediante el examen documental de cartas, diarios íntimos, artículos de periódicos e informes oficiales, en muchas investigaciones se ha intentado desvelar con cierta sistematicidad y reproductibilidad qué contienen y qué implican no solo los productos orales y escritos de los seres humanos, sino también toda obra humana dotada de significado.

En este sentido, ¿qué es lo que puede hacerse con los textos contenidos en los mensajes de Twitter? Aquí presentamos diversas propuestas de muy distintos signos. En primer lugar, adoptando la postura berelsoniana (Berelson y Lazarsfeld 1947), el análisis puede limitarse a cifrar las menciones o etiquetas, como ya se hizo en el apartado anterior.

En segundo lugar, también puede hacerse un recuento más cauto de las palabras que aparecen y sus coapariciones. Este procedimiento de análisis de contenido de texto no se limita a ser una elaboración estadística de una extracción de palabras, sino que atraviesa por una serie de fases para tratar semánticamente los textos:

1. La construcción del corpus o conjunto de textos que se van a analizar
2. El procesamiento previo, que incluye entre otras operaciones, la radicalización, la lematización, la exclusión de palabras sin contenido (artículos, adjetivos determinantes, preposiciones y conjunciones) y la categorización.
3. El procesamiento posterior, que consiste en la eliminación de categorías en función de su alta o baja frecuencia.

De este tenor, una representación muy frecuente del recuento de vocablos es la denominada “nube de palabras”. Hay dos formas de realizarla: una a partir de las palabras y otra a partir de los lemas. En el gráfico 10 podemos observar una nube de palabras tradicional a partir de los mensajes originales en el contexto de #Cuéntalo. En contraste, el siguiente gráfico no contiene las palabras, sino los lemas: una representación de palabras que agrupa todas las formas verbales, o el género y número de los nombres y adjetivos.

Como las modalidades de las primeras formas gramaticales son mucho más numerosas que las restantes, el resultado suele ser que en la segunda modalidad predominan los verbos, mientras que en las nubes de palabras tradicionales destacan nombres y adjetivos. Obviamente, no se encuentran artículos, adjetivos determinativos, preposiciones ni conjunciones, porque se ha incluido una lista de palabras ocultas (*stop-words*) que son descartadas de la lista por no tener un significado relevante.

Gráfico 10. Nube de palabras de los mensajes originales en el contexto de #Cuéntalo.



Gráfico 11. Nube de lemas de los mensajes originales en el contexto #Cuéntalo



FUENTE: Elaboración propia a partir de datos recogidos con *DiscoverText* (Shulman 2011).

En el primer gráfico destaca, además de la aparición de palabras vinculadas con el género (hombre, tío, mujer, amigo, amiga), una relacionada con la edad de las personas citadas en el mensaje (años) —tan frecuente porque son muchos los relatos de agresiones sexuales a menores— y diversas relacionadas con lugares (casa, coche, calle, fiesta). En el gráfico 11, el de lemas, destacan los verbos (haber, tener, decir, ir, pasar, salir, correr). Finalmente, ocupando el centro de ambos gráficos, destaca la palabra miedo, al estar contenida en muchos mensajes de este contexto.

La tercera propuesta de análisis de contenido pasa por la categorización, que añade una intencionalidad teórica a lo que se va buscando en el texto. La categorización puede hacerse de diversos modos, que van desde la construcción de un diccionario que automáticamente vaya clasificando todas las palabras en categorías, hasta la codificación manual, de acuerdo con una serie de reglas cuya correcta aplicación (intersubjetividad) se verifica mediante el empleo de medidas que evalúan el grado de concordancia de los codificadores en la aplicación de los códigos.

En el primer extremo se encuentra el llamado análisis de sentimiento (Blasco-Duatis y Coenders 2020). Este consiste en asignar a cada palabra de un texto una valencia positiva o negativa, sumándolas de modo que el conjunto arroje un valor positivo, negativo o neutro. Sin embargo, el análisis de sentimientos, aunque muy utilizado, no está exento de problemas. Esto es más notable cuando se analizan textos en castellano, ya que los diccionarios existentes suelen ser meras traducciones de otros en inglés. Esto se observa especialmente en los tuits: son textos muy cortos donde la ironía, el doble sentido, los iconos, los vídeos y las

imágenes que los acompañan no son contemplados adecuadamente con esta técnica. No obstante, podría servir para detectar algunos de los textos más agresivos o benignos.

Aplicando el diccionario *ncr* del paquete *syuzhet* de R (Jockers 2015) y recogiendo solo los dos mensajes con sentimientos más negativos y positivos del contexto de #Cuéntalo, se pueden comprender las inexactitudes de este procedimiento:

Imagen 2. Mensaje calificado como negativo de #Cuéntalo (anonimizado).

Supé que haciéndome daño a mí no iba a sentirme mejor, sino que entraría en la espiral que ya conocía de culpa y dolor y más daño que no resuelve nada. Y rompí con todo. Rompí con todo aquel que le molestaba verme triste, mal. Pero yo necesitaba estarlo. #cuéntalo



FUENTE: Twitter.

Este mensaje se valora como negativo al ser largo y contener muchas palabras con carga determinada como negativa (*daño, culpa, dolor, romper, molestar, triste, mal*). Por otro lado, el mensaje valorado con sentimiento más positivo es el siguiente:

Imagen 3. Mensaje calificado como positivo de #Cuéntalo (anonimizado).

Salir de fiesta y que un amigo tenga que acompañarte dos manzanas lejos de su casa, porque ir sola siendo mujer sigue sin ser seguro #Cuéntalo

11:38 p. m. · 28 abr. 2018 · Twitter for iPhone

1 Retweet 1 Tweet citado 2 Me gusta

FUENTE: Twitter.

Este otro mensaje no es tan largo como el anterior, y aunque su contenido sea más bien negativo, la aparición de palabras como *salir, fiesta, amigo, acompañar, seguir* e incluso *seguro* hace que el algoritmo lo considere positivo. Lo más erróneo en esta circunstancia es que la palabra *seguro* está precedida por “sin ser” y es valorada como positiva, cuando en realidad da la vuelta a todo el mensaje por muchas palabras positivas que contenga. De los millares de

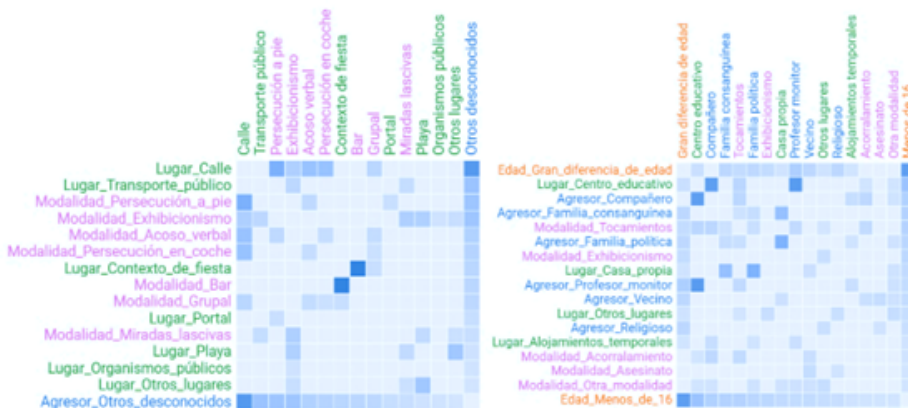
mensajes originales recogidos, el 31% se han evaluado como neutros (con una puntuación de 0), el 34% se han calificado como positivos y solo un 35% como negativos. Hay, pues, un equilibrio que no se corresponde con lo esperable del contexto en estudio, que por su propia finalidad debería ser negativo.

En el otro extremo de la tercera propuesta, se encuentra lo que se denomina teoría fundamentada, que consiste en la construcción iterativa y consensuada del sistema de categorías con el que se pretende descubrir tanto significado de un texto junto como su contexto. Entre estos dos últimos sistemas de categorizar los textos, se puede ubicar una práctica consistente en realizar al texto una serie de preguntas con sus posibles categorías de respuestas elaboradas de un modo flexible de modo que las poco frecuentes en el análisis puedan ser combinadas, al tiempo que se puedan generar otras a medida que se va practicando la categorización.

Un ejemplo de este proceso puede hacerse con los datos con la etiqueta de #Cuéntalo. Para ello se realiza en primer lugar una división entre aquellos mensajes que contienen relatos de violencia sexual y aquellos que no. Obviamente, dado el contexto de este conjunto de mensajes, solo los primeros son de interés. Una vez seleccionados, nos podríamos hacer las siguientes preguntas: ¿quién o quiénes fueron los agresores?, ¿qué edad tenía la víctima o víctimas?, ¿de qué tipo de agresiones se trataba? y ¿dónde se produjo este tipo de conducta? Este proceso de codificación ha de hacerse en buena parte manualmente, pero ya existen herramientas que pueden simplificar el proceso a través de la construcción de un diccionario en el que para cada categoría se especifica una serie de palabras que dan pistas de que al texto en cuestión se le aplica el código correspondiente.

Aplicado al contexto en estudio, esas cuatro preguntas dieron lugar a 49 categorías distintas: 11 tipos de agresores, 19 modalidades de acoso, 3 edades y 16 lugares distintos. Las categorías o respuestas más frecuentes a los cuatro interrogantes fueron respectivamente *agresores desconocidos* (53%), *tocamientos* y *acoso verbal* (ambas con 32%), *menores de 16 años* (29%) y *en la calle* (35%). Como se muestra en los mapas de colores de las categorías “*Otros desconocidos*” y “*Menores de 16 años*” (última columna a la derecha de ambas matrices), las asociaciones más fuertes (situadas en las primeras líneas de las matrices) en los episodios compartidos en la red relacionados con “*Otros desconocidos*” son haber transcurrido en la calle o en el transporte público, consistente en persecuciones a pie o coche, exhibicionismos o acosos verbales. En el caso de los episodios que ocurrieron cuando la víctima tenía “*Menos de 16 años*”, lo más común es que el agresor tuviera una gran diferencia de edad, que la acción tuviera lugar en un centro educativo o en la propia casa y que los agresores fueran un familiar consanguíneo o político, un profesor o un vecino. En este caso las agresiones denunciadas más comunes fueron los tocamientos y los exhibicionismos.

Gráfico 12. Mapas de color con las asociaciones más frecuentes de la categoría Otros desconocidos (izquierda) y Víctima con menos de 16 años (derecha).



FUENTE: Elaboración propia a partir de datos recogidos con *Discovertext* (Shulman 2011).

Finalmente, en la cuarta propuesta, hay que referirse a los modelos de clasificación automáticos. Estos proceden de lo que ha venido llamándose el aprendizaje automático (Molina y Garip 2019), al que se le ha sumado el denominado procesamiento de lenguaje natural y ha dado lugar al aprendizaje profundo. Sin ir demasiado lejos en lo que estas técnicas pueden contribuir a un análisis de los textos, es preciso mencionar que existen dos grandes tipos de procedimientos: los análisis supervisados y los análisis no supervisados.

En los primeros es preciso dividir el universo de textos que se desea analizar en una muestra de ejecución y otra de aplicación. Sobre la primera se realiza una clasificación manual en la que se aplica un algoritmo para la detección de palabras que pueden haber sido relevantes para los que hayan realizado la clasificación. Una vez que se logra un modelo que prediga bien lo que los codificadores ya han hecho, se aplica el modelo a la segunda muestra para realizar la clasificación oportuna.

En el contexto de #Cuéntalo, de los más de 32.000 mensajes originales que se recabaron con la herramienta de *Discovertext*, se hizo una muestra aproximada de 10.000. Sobre ellos un par de codificadores los clasificaron según tuvieran o no un contenido de acoso sexual. Con esta información se obtuvo un modelo basado en la frecuencia de aparición de las palabras, que se aplicó al corpus de más de 50.000 mensajes originales, a partir del cual se clasificó el conjunto total en estas dos categorías con un determinado grado de certidumbre. El algoritmo utilizado fue el de *WordStat* para clasificación supervisada. El resultado fue un 56% de mensajes sin agresiones, 12% de indefinidos y 32% de mensajes con agresiones. Este resultado indica que la muestra de 10.000 mensajes fue realizada aleatoriamente y que la clasificación supervisada da muy buenos resultados.

En el segundo tipo de análisis se encuentran los métodos no supervisados, entre los que destaca el algoritmo LDA para análisis de texto (Blei *et al.* 2003). De modo similar a lo que realiza el conocido método clasificatorio K-means, el investigador ha de proporcionar un número de temas en los que se quiere clasificar un conjunto de textos. De hecho, todos los mensajes disponibles se clasificaron en 20 áreas temáticas de las que en la tabla 9 se exponen solo las cuatro en las que se clasificaron más de 1000 mensajes.

La temática más numerosa (7% de los mensajes) corresponde, dadas las palabras y los *ngramas* característicos de ella (tabla 9), a aquellas menciones sobre los peligros de volver a casa sola (casa, llaves, sola, volver a casa). El segundo (que incluye casi un 5% de mensajes) corresponde a escenas del colegio cuando era pequeña (años, tenía, colegio). El tercer tema, compuesto por un 2,5% de mensajes, comprende una parte de los mensajes en catalán que se escribieron con la etiqueta de #Cuéntalo, pues todas las palabras características son catalanas (anys, vaig, em vaig). Ello refleja que la cuestión traspasó la frontera lingüística en una comunidad con una gran experiencia en la participación colectiva en las redes, como se puso de manifiesto el año anterior a la sentencia durante el Pro-cés. Finalmente, el cuarto (2%), podría reconocerse como el identitario al tener en su contenido la palabra mujeres, la locución “todas las mujeres” y muchos verbos en primera persona del plural.

Con este elenco de posibles análisis de contenido no se agotan los modos de aproximarse a los mensajes emitidos en Twitter. A este respecto, si se desea un examen de corte más ideológico o político, sería conveniente aproximarse a una literatura más especializada como es el caso de Alonso *et al.* (2012) o Grimer y Stewart (2013).

Tabla 9. Principales temas obtenidos mediante clasificación automática.

# Tema	Palabras Clave	% Casos	Coherencia
1 Sola a Casa Llaves	Casa; Llaves; Sola; Noche; Volver; Miedo; Sola a Casa; Volver a Casa; Casa Sola; Casa de Noche; Volviendo a Casa; Volver Sola;	7.16%	0.384
2 Años Tenía	Años; Tenía; Colegio; Pequeña; Niños; Era;	4.70%	0.328
3 Em Va Anys	Em; Anys; Amb; Vaig; Els; Va; Ho; El; És; Em Va; Em Vaig;	2.43%	0.523
4 Todas las Mujeres	Todas; Mujeres; Estamos; Tenemos; Somos; Hemos; Historias; Muchas; Todas las Mujeres; Todas Hemos; Hemos Sufrido; Todas Tenemos; Hemos Vivido; Muchas Mujeres; Estamos Solas;	1.99%	0.341

Fuente: Elaboración propia a partir de datos recogidos con T-Hoarder (Congosto *et al.* 2017).

9 CONCLUSIONES.

A lo largo de este artículo se han presentado diversos procedimientos para obtener y analizar la información disponible en Twitter. Se han dividido estos análisis en cuatro tipos: dos primeros muy relacionados, el análisis descriptivo y el dinámico, un tercer tipo muy propio del análisis de redes sociales, que sería el análisis estructural, y un cuarto, más próximo a metodologías cualitativas, que es el análisis de contenido.

Aplicando estos cuatro análisis a los mensajes escritos con el *hashtag* #Cuéntalo durante los días 26 y 28 de abril de 2018, se ha mostrado la utilidad de las técnicas empleadas. De esta forma, se ha destacado que este episodio de narrativa virtual tuvo un alto grado de viralidad, llegando en tres días a circular decenas de miles de mensajes distintos. Mediante el análisis de los perfiles participantes se ha revelado que @LaFallaras fue la usuaria de la red más activa desde el inicio y que progresivamente se fueron uniendo a la denuncia más y más cuentas compartiendo y publicando nuevos contenidos. Así, gracias a la participación de escritoras, periodistas y tuiters populares, y a la promoción llevada a cabo por asociaciones, periódicos y políticos, se desarrolló un gran movimiento en contra de la sentencia de La Manada como se ha podido comprobar a través de los *hashtags* asociados. El mensaje más difundido, firmado por @martolius, es un claro ejemplo del tipo de contenido de la reivindicación: son mensajes de denuncia, contando situaciones de violencia sexual y machismo padecidas por las mujeres. A través del análisis de contenido, sabemos que los principales temas versaron el acoso callejero, el miedo a volver a casa sola y agresiones ocurridas cuando eran menores de 16 años, que es la edad de consentimiento sexual en España.

Con un evidente prisma metodológico este artículo ha presentado un esquema de explotación de la información presente en las redes sociales para lo que ha explorado los modos de selección, los objetos de análisis posibles y las herramientas analíticas más centrales. Aunque por falta de espacio no se hayan tratado con mayor detalle, ni se hayan cubierto todas sus posibilidades y limitaciones, las propuestas, aquí enfocadas en Twitter, son fructíferas y pueden ser aplicadas a otras plataformas como Facebook, YouTube, Instagram o TikTok, para lo que sería conveniente también adentrarse en el análisis de imágenes y vídeos. El constante aumento del uso y popularidad de las redes sociales y su consecuente efecto en nuestra vida deja en evidencia la importancia de analizar más y mejor la interacción humana con y en la red. El papel de los científicos sociales aquí es crucial, por lo que deberíamos ser capaces de captar, tratar e interpretar en profundidad estas nuevas formas masivas de comunicación.

10 BIBLIOGRAFÍA

- ALONSO, S., VOLKENS, A. y GÓMEZ, B. (2012). *Análisis de contenido de textos políticos. Un enfoque cuantitativo*, Madrid, CIS.
- ARCILA, C., BLANCO, D. y VALDEZ, M. B. (2020). “Rechazo y discurso de odio en Twitter: análisis de contenido de los tuits sobre migrantes y refugiados en español”. *Revista Española de Investigaciones Sociológicas* (172), 21-40.
- BERELSON, B. y LASZARFELD, P. (1947): *The Analysis in Communications Content*. Nueva York, University of Chicago y Columbia University.
- BLASCO-DUATIS, M. y COENDERS, G. (2020). “Análisis de sentimiento de la agenda de los partidos políticos españoles en Twitter durante la Moción de Censura de 2018: Un enfoque de datos composicionales”. *Revista Mediterránea de Comunicación*, 11(2), 185-198.
- BLEI, D.M., NG, A.Y., y JORDAN, M.I. (2003). *Latent Dirichlet Allocation*. *The Journal of Machine Learning Research*, 3, 993-1022.
- BLONDEL, V. D., GUILLAUME, J. L., LAMBIOTTE, R., & LEFEBVRE, E. (2008). “Fast Unfolding of Communities in Large Networks”, *Journal of Statistical Mechanics*, P10008.
- BOVET, A. y MAKSE, H. (2019). “Influence of Fake News in Twitter during the 2016 US presidential election”. *Nature Communications*, 10(1), 1-14.
- BUCALO, M. S., CALVO, L., CUCCHIETTI, F., GARCÍA POVEDANO, D., GARCÍA-SÁEZ, A., FELIPE GÓMEZ, J. & FERNANDA VÉLEZ, D. (2019). “A Constellation of Horrors: Analysis and Visualization of the #Cuéntalo Movement”. *Companion Proceedings of The 2019 World Wide Web Conference*, 751-754.
- CASERO-RIPOLLÉS, A. (2020). “Influence of Media on the Political Conversation on Twitter: Activity, Popularity, and Authority in the Digital Debate in Spain”. *Icono14*, 18(1), 33-57.
- CONGOSTO, M.L. (2018). Contando cómo se difundió el #cuéntalo, disponible en <http://www.barriblog.com/2018/05/contando-se-difundio-cuentalo/> [consulta: 9/12/2020].
- CONGOSTO, M.L., BASANTA-VAL, P. y SÁNCHEZ-FERNÁNDEZ, L. (2017). “T-Hoarder: A Framework to Process Twitter Data Streams”. *Journal of Network and Computer Applications*, 83, 28-39.
- DEAL, B. E., MARTINEZ, L. S., SPITZBERG, B. H., & TSOU, M. H. (2020). “«I Definitely Did Not Report It When I Was Raped... #WeBelieveChristine #MeToo»: A Content Analysis of Disclosures of Sexual Assault on Twitter”. *Social Media+ Society*, 6(4), 2056305120974610.
- ESCOBAR M. y MARTÍNEZ-URIBE L (2020). “Network Coincidence Analysis: The netCoin R Package”. *Journal of Statistical Software*, 93(11), 1–32.
- FREELON, D., McILWAIN, C. y CLARK, M (2016). “Quantifying the Power and Consequences of Social Media Protest”, *New Media & Society*, 20(3), 990-1011.
- GILL, R. y ORGAD, S. (2018). “The Shifting Terrain of Sex and Power: From the ‘Sexualization of Culture’ to #MeToo”. *Sexualities*, 21(8), 1313-1324.
- GLEASON, B. (2013). “#Occupy Wall Street: Exploring Informal Learning About a Social Movement on Twitter”. *American Behavioral Scientist*, 57(7), 966-982.
- GONZÁLEZ-BAILÓN, S., BORGE-HOLTHOEFER, J. y MORENO, Y. (2013). “Broadcasters and Hidden Influentials in Online Protest Diffusion”. *American Behavioral Scientist*, 57(7), 943-965.
- GUALDA, E. y REBOLLO, C. (2020). “Big data y Twitter para el estudio de procesos migratorios: Métodos, técnicas de investigación y software”. *Empiria: Revista de*

- Metodología de Ciencias Sociales* (46), 147-177.
- GUTIÉRREZ, M., PANDO, M.J. y CONGOSTO, M. (2020). "New Approaches to the Propagation of the Antifeminist backlash on Twitter". *Investigaciones Feministas*, 11(2), 221-237.
- GRIMMER, J., & STEWART, B. M. (2013). "Text as data: The promise and pitfalls of automatic content analysis methods for political texts". *Political analysis*, 21(3), 267-297.
- JACKSON, S. J., BAILEY, M., & WELLES, B.F. (2018). "#GirlsLikeUs: Trans advocacy and community building online". *New Media & Society*, 20(5), 1868-1888
- JACKSON, S. J., BAILEY M. y WELLES, B. F. (2020). *"#HashtagActivism: Networks of Race and Gender Justice"*. Cambridge: The MIT Press.
- JACOMY, M., VENTURINI, T., HEYMANN, S., & BASTIAN, M. (2014). "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software". *PLoS ONE* 9(6): e98679.
- JENSEN, M. y BANG, H. (2013). "Occupy Wall Street: A New Political Form of Movement and Community?", *Journal of Information, Technology and Politics*, 10(4), 444-461.
- JOCKERS, M.L. (2015). "Syuzhet: Extract Sentiment and Plot Arcs from Text", disponible en <https://github.com/mjockers/syuzhet>, accedido el 10/12/2020.
- JOST, J. et al. (2018): "How Social Media Facilitates Political Protest: Information, Motivation, and Social Networks". *Political Psychology*, 39(1), 85-118.
- KEARNEY, M.W. (2019). "rtweet: Collecting and Analyzing Twitter Data." *Journal of Open Source Software*, 4(42), 1829. DOI: 10.21105/joss.01829, R package version 0.7.0.
- KUNST, J. R., BAILEY, A., PRENDERGAST, C., & GUNDERSEN, A. (2019). "Sexism, rape myths and feminist identification explain gender differences in attitudes toward the #metoo social media campaign in two countries". *Media Psychology*, 22(5), 818-843.
- LARRONDO, A., MORALES-I-GRAS, J., & ORBEGOZO-TERRADILLOS, J. (2019). "Feminist hashtag activism in Spain: Measuring the degree of politicisation of online discourse on #YoSíTeCreo, #HermanaYoSíTeCreo, #Cuéntalo y #NoEstásSola". *Communication & Society*, 207-221.
- LARSSON, A. y MOE, H. (2014). "Triumph of the Underdogs? Comparing Twitter Use by Political Actors during two Norwegian Election Campaigns". *Sage Open*, 4(4), 1-13.
- MOLINA, M. y GARIP, F. (2019). "Machine Learning for Sociology". *Annual Review of Sociology*, 45, 27-45.
- NAVARRO, C. y COROMINA, Ò. (2020). "Discussion and Mediation of Social Outrage on Twitter: The Reaction to the Judicial Sentence of «La Manada»". *Comunicación y Sociedad*, 33(1), 93-106.
- PÉREZ-DASILVA, J., MESO, K. y MENDIGUREN, T. (2020). "Fake news y coronavirus: detección de los principales actores y tendencias a través del análisis de las conversaciones en Twitter". *El Profesional de la Información*, 29(3), e290308.
- PÉREZ-MARTÍNEZ, V.M. y RODRÍGUEZ-GONZÁLEZ, MD (2017). "Movilización y participación en Twitter. Estudio de caso del hashtag #SuperTuesday en las primarias presidenciales de EEUU 2016". *Revista Latina de Comunicación Social*, 72, 679 a 703.
- PORTER, M.A.; ONNELA, J.P. Y MUCHA, P.J. (2009). "Communities in Networks". *Notices of the American Mathematical Society*. 56: 1082-1097.

- RODINO-COLOCINO, M. (2018). "Metoo, #MeToo: Countering Cruelty with Empathy", *Communication and Critical/Cultural Studies*, 15(1), 96-100.
- RUIZ, V. y VALLE, M.A. (2020). "#Cuéntalo: the Path Between Archival Activism and the Social Archive (s)". *Archives and Manuscripts*, 1-20.
- SHAO, C. et al. (2018). "The Spread of Low-Credibility Content by Social Bots". *Nature Communications*, 9(1), 1-41.
- SHULMAN, S. (2011). "DiscoverText: Software Training to Unlock the Power of Text", *Conference: Proceedings of the 12th Annual International Conference on Digital Government Research*, DGO. College Park, MD, USA.
- THEOCHARIS, Y., LOWE, W., VAN DETH, J. W., & GARCÍA-ALBACETE, G. (2015). "Using Twitter to mobilize protest action: online mobilization patterns and action repertoires in the Occupy Wall Street, Indignados, and Aganaktismenoi movements". *Information, Communication & Society*, 18(2), 202-220.
- THOMAS, W. y ZNANIECKI, F. (2004: 1918). *El campesino polaco en Europa y en América*, Madrid, Centro de Investigaciones Sociológicas (CIS).
- WANG, C. y WANG, P. (2013). "Discussing Occupy Wall Street on Twitter: Longitudinal Network Analysis of Equality, Emotion, and Stability of Public Discussion". *Cyberpsychology, Behavior and Social Networking*, 16(9), 679-685.
- XIONG, Y., CHO, M., & BOATWRIGHT, B. (2019). "Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement". *Public relations review*, 45(1), 10-23.

