

# Tendencias internacionales sobre análisis de redes sociales

Wenceslao Arroyo-Machado<sup>1</sup> (wences@ugr.es), Domingo Barroso-Hurtado<sup>1</sup> (dobarroso@ugr.es),  
Daniel Torres-Salinas<sup>1</sup> (torressalinas@ugr.es)

<sup>1</sup>Universidad de Granada

**Resumen:** La Web 2.0 ha dado lugar a numerosas plataformas en las que las interacciones entre sus usuarios generan grandes cantidades de datos, fácilmente accesibles. Entre las metodologías empleadas para analizar estos datos, el análisis de redes sociales se ha convertido en una de las más populares, aplicándose con éxito a campos de investigación muy diversos. En este capítulo introducimos algunos de los conceptos fundamentales en el análisis de redes sociales, enfocado al ámbito de la educomunicación. Todos los elementos que intervienen en este son puestos así en relieve, desde la planificación y recolección de datos a la generación y manipulación de la red, repasando las principales herramientas, indicadores y métodos que en él intervienen. Por último, ejemplificamos todo el proceso descrito con un estudio de caso relativo al debate producido en redes en torno a la nueva ley de reforma educativa.

**Zenodo:** <https://doi.org/10.5281/zenodo.5047157>

## 1. Introducción

Las tecnologías de la información y la comunicación están cada vez más presentes en nuestras vidas. Encontramos un mundo en el que no sólo hemos de lidiar con el uso de una gama más amplia de dispositivos y medios tecnológicos, sino en el que nuestra vida se ven afectada por su influencia y desarrollo. La evolución y el uso de las TICs (Tecnologías de la Información y Comunicación), y especialmente la incidencia de los medios sociales, entonces, habrían abierto la posibilidad a nuevas formas de comunicación, diálogo, interacción, creación de significado o participación social (Gutiérrez et al., 2018). Las corrientes y teorías relacionadas con la educomunicación se han preocupado por el papel e incidencia de los medios de comunicación y, posteriormente, de las TICs, en nuestra vida. Lo han hecho, al menos, desde su institucionalización desde los años 70, con el interés de organizaciones internacionales como la UNESCO (Barbas, 2012).

La educomunicación es un campo de estudio que encuentra una gran diversidad de términos asociados, como los de alfabetización mediática o educación mediática (Gutiérrez et al., 2018). También es diversa en cuanto a las perspectivas desde las que se desarrolla. En este sentido, puede ser entendida o bien en términos instrumentales o en términos más críticos (por ejemplo dialógicos).

Desde la primera perspectiva, hace referencia al uso y las habilidades para utilizar los medios y las TICs. Desde la segunda, hace referencia a una vertiente más transformadora, basada en el diálogo y la creación y construcción colectiva, así como en la comprensión del rol que juegan los medios de comunicación y las TICs a la hora de transformar nuestras relaciones, nuestra vida social o, incluso, cómo funcionan las estructuras de poder en las que los medios se insertan y a las que contribuyen (Barbas, 2012).

El análisis de redes (*network analysis*) representa un esfuerzo teórico-metodológico que nos puede ayudar a analizar y comprender mejor estos cambios en las estructuras de relaciones. Puede ayudarnos, por ejemplo, a comprender las relaciones con diferente forma, dirección e intensidad presentes entre diferentes actores y que pueden dar pie a diferentes relaciones de poder, e influir y abrir diferentes espacios para el ejercicio de su agencia (Rodríguez, 1995). El análisis de redes sociales (*social network analysis*) tendría una preocupación similar, pero centrada en las estructuras sociales (Marin y Wellman, 2014), inclusive los medios sociales y la Web 2.0 (Snee, 2008).

Existe una multiplicidad de estudios que utilizan datos de medios sociales (Sloan y Quan-Haase, 2017) y que han sido abordados positivamente en diferentes especialidades (Myers et al., 2014; Riquelme y González-Cantergiani, 2016; Robinson-García et al., 2018). Sin embargo los que más proliferan son los relativos a Facebook y Twitter (Ledford, 2020), si bien no faltan otros como Wikipedia, Instagram, Reddit o noticias de medios de comunicación (McCay-Peet y Quan-Haase, 2017). Es por ello que el análisis de redes sociales, en especial aplicado a datos de Twitter, debido a su carácter abierto y diferentes tipos de interacciones (tuits, retuits, me gusta y relaciones de seguimiento), se ha convertido en una relevante fuente para mapear la actividad y relaciones de una comunidad concreta (Darmon et al., 2015). Específicamente, dentro de las ciencias sociales y humanidades, se ha encontrado en esta metodología un floreciente campo de investigación aplicado a áreas tan diversas como la red global de humanidades digitales (Grandjean, 2016), la comunicación electoral (Davis et al., 2017) o los estudios de la actividad científica (Robinson-García et al., 2019).

Sin embargo, como exponen Menashy y Verger (2019), el análisis de redes sociales todavía presenta un recorrido incipiente en campos del ámbito educativo y político, como los de la política educativa global y la educación internacional. No obstante, es posible encontrar algunos estudios que exploran debates en Twitter en materia de política social y educativa. Podemos enumerar algunas investigaciones, como la de Schuster, Jörgens y Kolleck (2019), que exploraron el debate y las interacciones entre actores en relación a la educación inclusiva, o la de Goritz, Kolleck y Jörgens (2019), que remitieron a los actores con una influencia más elevada en discusiones educativas sobre la sostenibilidad. Desde una perspectiva educomunicativa crítica, suponen, en definitiva, trabajos que pueden alimentar análisis sobre el uso de las redes sociales, las TICs y, por ejemplo, las relaciones de poder que pueden reflejar las relaciones entre actores en el seno de las mismas.

A lo largo de este capítulo introducimos algunos de los conceptos fundamentales en el análisis de redes sociales. Nos centramos en todos los procesos que participan en esta metodología, desde la recolección de datos a la generación e interpretación de la red. Finalizamos con un ejemplo relativo al debate producido en redes en torno a la nueva ley de reforma educativa. De esta manera, buscamos que el lector encuentre una guía intuitiva y accesible para iniciarse en este tipo de análisis y para comprender cómo este puede nutrir los análisis en el ámbito de la educomunicación.

## 2. Metodología

### 2.1. Materiales

#### 2.1.1. Los datos de medios sociales

Uno de los aspectos más relevantes del análisis de redes sociales se encuentra en su capacidad a la hora de poderse aplicar sobre cualquier estructura o fenómeno que responda a un sistema complejo (Ladyman et al., 2013): los actores que participan en este sistema y las relaciones entre ellos son ideales para representarse a través de una red. Por ello, uno de los entornos que mejor se adaptan a estos tipos de análisis lo encontramos en los medios sociales. Desde su nacimiento con la Web 2.0, los medios sociales no han dejado de crecer, dando lugar a todo un gran ecosistema de plataformas. Prueba de su volumen lo tenemos en los 2498 millones de usuarios activos en Facebook, 1000 millones en Instagram, 386 millones en Twitter, todos ellos a fecha de abril de 2020<sup>1</sup>, o 53.617.358 entradas en Wikipedia a 8 de junio de 2020<sup>2</sup>. Es por ello que estas se han convertido en grandes fuentes de las que extraer todo tipo de datos, siendo este volumen junto a su alto grado de interacciones sus puntos fuertes (Liu et al., 2016). Pero no están exentas de problemas y limitaciones. En la actualidad, los propios datos plantean tres grandes desafíos: su calidad, su dependencia respecto a sus proveedores y, por último, la heterogeneidad que los datos pueden presentar al combinar diferentes plataformas que arrojan datos de diferente tipo (Haustein, 2016). Se plantean así varios desafíos para la generación de resultados de investigación significativos, siendo de especial relevancia la identificación de temas y eventos en la conversación llevada a cabo, la obtención de datos de calidad y la correcta selección de herramientas y técnicas de visualización (Stieglitz et al., 2018).

Motivados por ello, los análisis basados en medios sociales han presenciado un cambio ocasionado por el aumento desmesurado de datos que producen, quedando las técnicas tradicionales para su tratamiento obsoletas y surgiendo así el denominado *social big data* o *big social data* (Bello-Organ et al., 2016). Los procesos involucrados en estos estudios, desde el almacenamiento y recolección de los datos a las técnicas de análisis y posterior interpretación de los resultados obtenidos, plantean ahora otros nuevos retos que afrontar desde las tecnologías *big data* (Sebei et al., 2018). Existe así un nuevo escenario que exige esfuerzos renovados para adaptar las técnicas usadas hasta ahora y hacer frente a nuevos problemas, pero que al mismo tiempo permite explorar otros horizontes.

De esta forma, cuando se afronta un estudio de redes sociales en contextos digitales, el primer punto a abordar es la recolección de datos, que plantea diferentes problemas en función del volumen y tipo de datos. En la actualidad, las APIs (*Application Programming Interface*) son la principal herramienta en la recopilación de datos, pues ofrecen el punto de acceso en la mayoría de las ocasiones a los datos brutos. Las APIs permiten simplificar el trabajo de los desarrolladores al ofrecer un conjunto de funciones y procedimientos específicos orientados a facilitar la captura y el procesamiento de los datos. Por ejemplo, en el caso de los medios sociales, están ampliamente extendidas para la recuperación de datos y la respuesta ofrecida suele ser texto en formato estructurado, usualmente JSON o XML. Gracias a ello no solo es posible alcanzar una mayor eficiencia en el proceso de recolección, sino que, además, se posibilita el acceso a metadatos que a menudo no se muestran públicamente. No obstante, como aspectos negativos, es necesario remarcar que no todas las plataformas cuentan con una API abierta a todo el público, siendo los datos ofrecidos y su acceso limitados (Littman et al., 2018; Thomson, 2016). Entre los medios que cuentan con API destacan Twitter, Facebook, Instagram, Reddit, Wikipedia y Youtube. Pero las APIs no son el único método de recolección, existiendo para ello otras alternativas como el uso de servicios ofrecidos por terceros,

<sup>1</sup> <https://datareportal.com/social-media-users> [Consultado el 8 de junio de 2020]

<sup>2</sup> [https://en.wikipedia.org/wiki/Wikipedia:Size\\_of\\_Wikipedia](https://en.wikipedia.org/wiki/Wikipedia:Size_of_Wikipedia) [Consultado el 8 de junio de 2020]

como los agregadores de datos que combinan contenidos de varios medios en una única plataforma, volcados de datos proporcionados por la plataforma o herramientas de *web scraping*. Estas últimas resultan de gran interés ya que permiten realizar una recuperación automatizada de información simulando el comportamiento humano, aunque por ello plantean una serie de cuestiones jurídicas y éticas que tienen que considerarse antes de realizar la recolección (Krotov y Silva, 2018).

Teniendo en cuenta las posibilidades planteadas a la hora de recolectar los datos, es necesario establecer una estrategia antes de hacerlo. En primer lugar, exceptuando algunos casos, los datos se obtienen tras la formulación de una pregunta de investigación y no al revés, algo en lo que el *big data* ha influido al facilitar el acceso a grandes cantidades de datos (Schroeder, 2014). Una vez establecida es necesario determinar: I. la cantidad de medios que vamos a emplear en el estudio: ¿queremos usar uno o varios medios?; II. el tipo de datos que necesitamos (interacciones entre usuarios, entre hashtags y/o términos o un combinación de ambos): ¿qué actores y relaciones queremos mapear?; III. el volumen de datos: ¿cuántos actores y relaciones van a ser estudiados? ; IV. el sesgo introducido en los datos: ¿son suficientemente representativos nuestros datos? (Mayr y Weller, 2016).

### 2.1.2. La importancia en el preprocesamiento de los datos

Una vez definida la estrategia de recolección, es necesario preparar los datos para trabajar con ellos y generar correctamente las redes. En primer lugar, hay que tener en cuenta que, en caso de haber recuperado datos de diferentes fuentes, el punto de partida está en combinarlos correctamente. Por ejemplo, al trabajar con menciones de hashtags que realizan los usuarios de Twitter y Facebook es necesario plantear cómo han de fusionarse ambos, no solo reduciendo los dos conjuntos de datos a los campos necesarios y comunes, como el usuario que realiza la mención y el hashtag, sino diferenciando el medio para que no se lleven a cabo solapamientos añadiendo un campo adicional que indique el medio desde el que se realiza. Se trata en definitiva de revisar cada uno de los conjuntos recuperados e integrarlos en una base de datos o plataforma para trabajar con ellos. Entornos como RStudio (para el lenguaje de programación R), Spyder (Python), Microsoft Excel, en aquellos casos que el volumen no sea muy extenso, o incluso gestores de bases de datos SQL (como MySQL, MariaDB o PostgreSQL) o NoSQL (MongoDB) permiten una ágil manipulación de los datos, pudiendo integrar conjuntos con diferentes estructuras y codificaciones, los dos principales problemas de esta fase.

El siguiente paso está en la limpieza y normalización de los datos. Una vez los hemos integrado en un mismo entorno, es preciso revisar la consistencia de todos ellos para evitar conflictos en el análisis de redes, corrigiendo problemas como valores anómalos o perdidos. Es un proceso imprescindible en la minería de datos y que puede requerir de técnicas *big data* en aquellos casos en los que se trabaje con grandes volúmenes de datos (García et al., 2016). Sobre todo al trabajar con conjuntos de datos ofrecidos por terceros, es posible que algunos campos no aparezcan, como el nombre de usuario, su localización o fecha de mención, pudiendo ese problema deberse a la inexistencia de dicho valor o un problema en la recopilación efectuada. Además, en algunos casos se pueden encontrar campos cuyos valores aparezcan codificados, como al usar datos de Twitter ofrecidos por terceros, ya que su política establece que han de compartirse usando únicamente el identificador de usuario y tuit<sup>3</sup>. Es por ello que en casos como el descrito, sería necesario acudir a la API usando los identificadores para descifrar dichos campos e integrarlos en nuestros datos.

---

<sup>3</sup> <https://developer.twitter.com/en/developer-terms/agreement-and-policy>

Tenemos así una amplia gama de medios sociales de los que es posible establecer diferentes tipos de relaciones. Por un lado están las redes creadas a partir de las interacciones generadas entre sí por los diferentes actores sociales, como pueden ser las menciones o relaciones de seguimiento o amistad (Shields, 2016), las cuales se dan entre diferentes usuarios, como en Twitter, Facebook o Instagram. Por otro lado, se encuentran las relaciones entre otra entidades, como las menciones a artículos científicos o co-ocurrencias de palabras y hashtags (Haunschild et al., 2019). En último lugar, es posible establecer un tipo de red mixta que combine ambas anteriores, teniendo por ejemplo redes de actores sociales que citan trabajos científicos o mencionan hashtags (Robinson-García et al., 2019). Estas redes, además, no solo se suscriben a un único medio social, siendo posible combinar varios de ellos. Más allá de los propios medios sociales, es posible desarrollar un análisis de redes sociales en plataformas educativas atendiendo a las diferentes tipologías de interacciones antes señaladas (Giri et al., 2014; Rabbany et al., 2011; Zhuhadar y Butterfield, 2014).

### 2.1.3. Generación del grafo

Una vez que los datos han sido preparados estos pueden usarse para generar una red. Al respecto, es necesario aclarar que si bien el análisis de redes sociales permite modelar sistemas complejos y centra su atención en las conexiones que los diferentes actores establecen y el papel que desempeñan, es la teoría de grafos la que ofrece una estructura matemática subyacente a la red y realiza con ello una abstracción de la misma. Es decir, mientras que la teoría de grafos establece la estructura gráfica y ofrece una serie de indicadores y métricas para estudiarlos, el análisis de redes permite realizar deducciones sobre el sistema que se está estudiando. En lo referente a los grafos, estos están compuestos principalmente de dos elementos: los nodos y las aristas. Mientras que los primeros representan los actores, los segundos hacen lo propio con las interacciones o relaciones que estos tienen entre sí, permitiendo establecer caminos entre pares de nodos. Si bien existen diferentes tipos de grafos en función de las aristas, en este caso los dos predominantes son los dirigidos —las aristas reflejan direccionalidad— y los no dirigidos —las aristas muestran una relación recíproca— (Jackson, 2010; West, 2000).

Existen múltiples formatos a partir de los cuales se pueden generar los grafos. Por un lado están los obtenidos a partir de un software, como los ficheros con la extensión NET (Pajek), mientras que otros tienen un carácter más universal, véase GML (*Geography Markup Language*), GEXF (*Graph Exchange XML Format*) o incluso archivos delimitado por comas, puntos y comas (CSV) o tabuladores (TSV). Con independencia de su lenguaje o, incluso, etiquetado en aquellos basados en XML, la estructura suele constar de dos partes a modo de tablas: nodos y aristas. Los nodos vienen establecidos de manera única, asociando cada uno a un identificador, y pueden contar con tantas características como sea necesario, mientras que las aristas aparecen como relaciones entre pares (nodo de origen y nodo de destino), empleando para ello los identificadores de los nodos, pudiendo fijar también características adicionales a cada arista como el peso de la misma. Estas similitudes y diferencias pueden apreciarse en la Figura 1.

Figura 1. Comparación de diferentes tipos de archivos de grafos.

|  |  |   |
|--|--|---|
| <pre> 1 &lt;?xml version="1.0" encoding="UTF-8"?&gt; 2 &lt;graph xmlns="http://www.gexf.net/1.3/" version="1.0" xmlns:xsi="http://www.gexf.net/1.3/xsd" 3 xsi:schemaLocation="http://www.gexf.net/1.3 http://www.gexf.net/1.3/gexf.xsd" 4 &lt;meta lastmodified="2009-07-07" /&gt; 5 &lt;creatorDepth 8.0/creator&gt; 6 &lt;description/&gt; 7 &lt;/meta&gt; 8 &lt;/graph defaultEdgetype="undirected" mode="static"&gt; 9 &lt;nodes&gt; 10 &lt;node id="1" label="Usuario 1"&gt; 11 &lt;viz:size value="10.0"/&gt; 12 &lt;/node&gt; 13 &lt;node id="2" label="Usuario 2"&gt; 14 &lt;viz:size value="10.0"/&gt; 15 &lt;/node&gt; 16 &lt;node id="3" label="Usuario 3"&gt; 17 &lt;viz:size value="10.0"/&gt; 18 &lt;/node&gt; 19 &lt;node id="4" label="Usuario 4"&gt; 20 &lt;viz:size value="10.0"/&gt; 21 &lt;/node&gt; 22 &lt;node id="5" label="Usuario 5"&gt; 23 &lt;viz:size value="10.0"/&gt; 24 &lt;/node&gt; 25 &lt;/nodes&gt; 26 &lt;edges&gt; 27 &lt;edge id="0" source="1" target="2"/&gt; 28 &lt;edge id="1" source="1" target="4"/&gt; 29 &lt;edge id="2" source="2" target="3"/&gt; 30 &lt;edge id="3" source="3" target="4"/&gt; 31 &lt;edge id="4" source="3" target="5"/&gt; 32 &lt;/edges&gt; 33 &lt;/graph&gt; </pre> | <pre> 1 *Vertices 5 2 1 "Usuario 1" 3 2 "Usuario 2" 4 3 "Usuario 3" 5 4 "Usuario 4" 6 5 "Usuario 5" 7 *Edges 8 1 2 9 1 4 10 2 3 11 3 4 12 3 5 </pre> | <pre> 1 Id;Label 2 1;"Usuario 1" 3 2;"Usuario 2" 4 3;"Usuario 3" 5 4;"Usuario 4" 6 5;"Usuario 5" 7 8 Source;Target 9 1;2 10 1;4 11 2;3 12 3;4 13 3;5 </pre> |
| Archivo GEXF   | Archivo NET  | Archivos CSV  |

Una vez generada la red es posible ir un paso más allá y transformar la estructura de la red para pasar de las relaciones directas entre los nodos a las de similitud conceptualmente establecidas entre ellos. Se trata de unos métodos populares en la bibliometría: co-citación y emparejamiento bibliográfico. La co-citación se produce cuando dos trabajos son citados por un mismo tercero o terceros, mientras que el emparejamiento se da cuando dos trabajos citan a un mismo tercero o terceros. Más allá de la bibliometría, estos métodos también son utilizados en análisis de redes, estando ambos adaptados a los medios sociales (Costas et al., 2017). De este modo, es posible llevar a cabo una abstracción de las relaciones directas entre nodos, dando paso a las co-relaciones subyacentes entre ellos en base a sus interacciones, como las co-menciones o emparejamientos producidos por las menciones de hashtags. Esto hace que sea más fácil identificar qué actores guardan una relación entre sí en base a su propia actividad o la de terceros.

## 2.2. Métodos

### 2.2.1. Métricas

Podemos encontrar un gran número y diversidad en lo referente a las métricas que se emplean en el análisis de redes. Destacan varias que permiten obtener una visión general de la red. La densidad ofrece el nivel de conectividad, realizando una ratio de conexiones que un nodo establece por todas las que puede establecer. El diámetro y el radio, mayor y menor distancia de entre todos los pares de nodos de la red, permiten conocer lo alejados o cercanos que están los nodos en sus casos más extremos. También encontramos la longitud de camino medio, promedio de los caminos más cortos entre todos los pares nodos, que indica la eficiencia de la red en cuanto a las conexiones.

Pero, de entre todas las medidas que existen para estudiar la red, resultan de especial importancia las de centralidad. A través de estos indicadores es posible determinar la relevancia que un nodo tiene dentro de la red y compararlo con el resto. Estos pueden dividirse en indicadores locales o globales. Los primeros muestran su importancia relativa respecto a su entorno y relaciones inmediatas, mientras que las segundas sí reflejan la relevancia que tienen dentro de la red (Benedictis et al., 2014). La Tabla 1 recoge los cuatro indicadores principales empleados comúnmente en la literatura científica (Das et al., 2018; Landherr et al., 2010). No existe un indicador único y válido para todos los problemas: todos ellos son complementarios y su selección radica en la naturaleza del análisis. Resulta contraproducente realizar, por lo tanto, el cálculo de cuantas medidas sea posible. En su lugar, debe evaluarse qué es necesario medir en la red, realizar una búsqueda de métricas y, teniendo en

cuenta la tipología de la red, ya que algunos indicadores no son válidos para todas, usar los que resulten verdaderamente significativos.

*Tabla 1. Principales indicadores de centralidad en análisis de redes.*

| Indicador                     | Tipo   | Descripción  |
|-------------------------------|--------|--|
| Centralidad de grado          | Local  | Número de conexiones que establece un nodo.  |
| Intermediación                | Global | Medida que refleja la frecuencia con la que un nodo forma parte de los caminos cortos entre el resto de nodos de la red.   |
| Cercanía                      | Global | Nivel de proximidad que un nodo tiene con respecto al resto de nodos de la red.  |
| Centralidad del vector propio | Global | Evolución de la centralidad de grado que muestra la influencia de un nodo en la red al no tener en cuenta solo el número de conexiones que establece sino la centralidad de estos. |

Fuente: elaboración propia

### 2.2.2. Detección de comunidades

Por otra parte, uno de los aspectos que más interés levanta del análisis de redes sociales se encuentra en la detección de comunidades. Con ello es posible determinar de forma automatizada grupos de actores que, en base a sus conexiones, guardan una relación en común. Un ejemplo de ello lo tenemos en las denominadas comunidades de interés, que muestran relaciones entre actores que prestan atención a un mismo tema, o la detección de tópicos en redes de co-ocurrencia de palabras (Robinson-García et al., 2019). Uno de los algoritmos más populares empleados para ello es el de Louvain (Blondel et al., 2008), el cual cuenta con variantes como el de Leiden (Traag et al., 2019). Ambas medidas ofrecen además un valor de modularidad ( $Q$ ), entre 0 y 1, que refleja la calidad de las comunidades detectadas. De manera heurística, se consideran significativas las comunidades cuando este se encuentra por encima de 0,3 (Chen et al., 2015).

### 2.2.3. Visualización

Un apartado muy relevante en el análisis de redes sociales está en la poda, cuyo objetivo es hacer la red más fácilmente interpretable, eliminando, por ejemplo, los nodos con un menor valor de grado o las aristas de menor peso. Esta modificación conlleva, sin embargo, una alteración en la propia topología de la red, algo que hay que tener en cuenta a la hora de analizarla. No obstante, existen otros elementos dentro de la visualización que no alteran la estructura de la red y solo repercuten en su representación. Por un lado, existen varios algoritmos a través de los cuales se puede realizar la distribución de los nodos en la red, como el popular ForceAtlas2 en Gephi, basado en fuerzas de atracción y repulsión entre nodos mediante las relaciones establecidas entre ellos (Jacomy et al., 2014). Y por otro, las redes permiten varias formas de representar los indicadores calculados. Es posible establecer los colores de los nodos en base a la comunidad que pertenecen o a una medida de centralidad, mientras que el tamaño de los propios nodos se puede fijar con otra de ellas, mostrando de esta forma claramente las diferencias existentes entre todos ellos.

### 2.2.4 Herramientas para el análisis de redes

Existen diferentes tipos de herramientas para llevar a cabo el análisis de redes. En la Tabla 2 se han reunido algunas de las más destacadas. Además, algunas de ellas, como Gephi y NodeXL, ofrecen opciones adicionales para obtener datos de medios sociales como Twitter.

Tabla 2. Principales herramientas para análisis de redes sociales.

| Nombre              | Plataformas                    | Descripción  | Última versión |
|---------------------|--------------------------------|--|----------------|
| <b>Aplicaciones</b> |                                |  |                |
| Gephi               | Multi                          | Herramienta open-source para la exploración y visualización de grafos.   | 09/2017        |
| Graphviz            | Multi                          | Herramienta open-source para la visualización de grafos.   | 06/2020        |
| Pajek               | Multi                          | Aplicación que cuenta con dos ediciones adicionales (PajekXXL y Pajek3XL) y permite el análisis y visualización de grafos.             | 06/2020        |
| SocioViz            | Web                            | Plataforma web que permite la recopilación de datos de Twitter y Facebook y análisis de redes.   |                |
| SocNetV             | Multi                          | Herramienta open-source para el análisis y visualización de grafos.  | 03/2019        |
| <b>Paquetes</b>     |                                |  |                |
| igraph              | R, Python, Mathematica y C/C++ | Biblioteca para la manipulación y análisis de grafos.  | 04/2020        |
| graph-tool          | Python                         | Biblioteca para la manipulación y análisis de grafos.  | 05/2020        |
| NetworkX            | Python                         | Biblioteca para la manipulación y estudio de redes complejas.  | 06/2020        |
| NodeXL              | Excel                          | Plugin para Microsoft Excel que permite realizar grafos. Además del modelo básico ofrece varias suscripciones con más funcionalidades. | 04/2020        |

Fuente: elaboración propia

### 2.3. Estudio de caso

En último lugar, se ha realizado un pequeño estudio de caso con el fin de ejemplificar todo el marco desarrollado al completo.

## 3. Resultados

### 3.1. Planificación

Nuestro problema de investigación se centra en la identificación de comunidades de usuarios en torno al debate producido en medios sociales sobre la ley de reforma educativa. En siguiente lugar se lleva a cabo una estrategia para la recolección de datos. En este caso se va a emplear un único medio, Twitter, y los datos a recopilar son los relativos a los hashtags. Se trabajará con un volumen no muy elevado, ya que la API de Twitter limita la actividad a los últimos siete días (del 17 al 24 de junio de 2020). La consulta realizada es la siguiente: (“ley de educación” OR LOMLOE OR @CelaaIsabel). En cuanto al sesgo de los datos, tenemos como limitación que en la consulta establecida para recuperar los tuits va a aparecer ruido debido a que hay tuits que mencionan la cuenta @CelaaIsabel pero no tienen relación con el tema. No obstante, son mucho más numerosos los tuits que entran dentro del campo de estudio que los que no y de no usar esa mención los datos se verían sustancialmente reducidos.



### *3.2. Recolección y preparación de los datos*

Una vez planteada la recolección, mediante Tweepy (Roesslein, 2020), biblioteca disponible en Python, se ha accedido a la API de Twitter y descargado todos los tuits posibles. De cara a facilitar el posterior tratamiento de los datos, la recuperación se ha hecho de manera estructurada, almacenándose directamente en archivos CSV. En total han sido dos los ficheros resultantes de la recolección. Por un lado se han descargado los tuits completos, identificando un tuit en cada línea del fichero con los siguientes campos: usuario que realiza el tuit, ID de usuario y texto completo del tuit. Mientras que en otro fichero se han recuperado las menciones de cada uno de ellos, siendo cada línea una mención con los siguientes campos: usuario que realiza la mención, ID de usuario, usuario mencionado.

Con RStudio y el paquete dplyr (Wickham et al., 2020) se ha llevado a cabo el tratamiento de los datos y revisado su consistencia. En total son 16.506 los tuits recuperados y de todos ellos 12.111 (73,37%) incluyen alguna mención a un cuenta de Twitter. Son 5647 los usuarios que realizan alguna de esta menciones, una media de 5,72 ( $\pm 23,11$ ) tuits por usuario. Tenemos de este modo 32.321 menciones a 2995 usuarios únicos. Sobre estas menciones se han generado la red de co-menciones, estableciendo aristas entre todos aquellos usuarios de Twitter que son co-mencionados por un mismo usuario. Cabe destacar que de la red se ha extraído a @CelaaIsabel ya que al ser uno de los términos de búsqueda y aparecer en buena parte de los tuits, ocupaba un papel demasiado central, alterando con ello el resto de relaciones. El resultado son dos archivos CSV, uno relativo a los nodos y otro con las aristas. En el caso de los nodos, se ha calculado e incluido el total de menciones que cada cuenta recibe y en el de las aristas el peso otorgado a las mismas depende de las veces que son co-mencionadas.

A continuación, en Gephi se han importado ambos ficheros normalizados con el objetivo de visualizar la red. Para obtener una representación óptima se han considerado solamente aquellos nodos que tienen un grado mínimo de 5, es decir son co-mencionados con al menos cinco usuarios diferentes. A continuación se ha obtenido la componente principal y, mediante el algoritmo de Louvain (Blondel et al., 2008), se han identificado un total de 11 comunidades, siendo el valor de modularidad ( $Q$ ) de 0,566, por lo que podemos determinar que éstas son significativas. Si se analizan los actores de cada una de las comunidades se observa cómo los usuarios se agrupan bajo diferentes temáticas y contextos. Por ejemplo, se observan comunidades con una clara orientación política, comunidades en torno a organismos institucionales o bien influencers dentro de la educación o la educación en Andalucía. Una vez que las comunidades se han dado por válidas, tanto por el valor de modularidad como por la coherencia temáticas de las mismas, se ha realizado la distribución de los nodos mediante ForceAtlas2, coloreando los nodos en base a las comunidades y asignando tamaño de acuerdo al total de menciones que reciben. La red resultante (Figura 2) está compuesta así de 435 nodos y 3632 aristas.



- challenges. *Information Fusion*, 28, 45-59. <https://doi.org/10.1016/j.inffus.2015.08.005>
- Benedictis, L. D., Nenci, S., Santoni, G., Tajoli, L., y Vicarelli, C. (2014). Network Analysis of World Trade using the BACI-CEPII Dataset. *Global Economy Journal*, 14(3-4), 287-343. <https://doi.org/10.1515/gej-2014-0032>
- Blondel, V. D., Guillaume, J.-L., Lambiotte, R., y Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10), P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Chen, Z., Jia, M., Yang, B., y Li, X. (2015). Detecting overlapping community in complex network based on node similarity. *Computer Science and Information Systems*, 12(2), 843-855. <https://doi.org/10.2298/CSIS141021029C>
- Costas, R., de Rijck, S., y Marres, N. (2017). Beyond the dependencies of altmetrics: Conceptualizing 'heterogeneous couplings' between social media and science. *The 2017 Altmetrics Workshop*. [http://altmetrics.org/wp-content/uploads/2017/09/altmetrics17\\_paper\\_4.pdf](http://altmetrics.org/wp-content/uploads/2017/09/altmetrics17_paper_4.pdf)
- Darmon, D., Omodei, E., y Garland, J. (2015). Followers Are Not Enough: A Multifaceted Approach to Community Detection in Online Social Networks. *PLOS ONE*, 10(8), e0134860. <https://doi.org/10.1371/journal.pone.0134860>
- Das, K., Samanta, S., y Pal, M. (2018). Study on centrality measures in social networks: A survey. *Social Network Analysis and Mining*, 8(1), 13. <https://doi.org/10.1007/s13278-018-0493-2>
- Davis, R., Bacha, C. H., y Just, M. R. (2017). *Twitter and elections around the world: Campaigning in 140 Characters or Less*. Routledge.
- García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., y Herrera, F. (2016). Big data preprocessing: Methods and prospects. *Big Data Analytics*, 1(1), 9. <https://doi.org/10.1186/s41044-016-0014-0>
- Giri, B. E., Manongga, D., y Iriani, A. (2014). Using social networking analysis (SNA) to analyze collaboration between students (case Study: Students of open University in Kupang). *International Journal of Computer Applications*, 85(1). <https://doi.org/10.5120/14808-3022>
- Goritz, A., Kolleck, N., y Jörgens, H. (2019). Education for Sustainable Development and Climate Change Education: The Potential of Social Network Analysis Based on Twitter Data. *Sustainability*, 11. <https://doi.org/10.3390/su11195499>
- Grandjean, M. (2016). A social network analysis of Twitter: Mapping the digital humanities community. *Cogent Arts & Humanities*, 3(1), 1171458. <https://doi.org/10.1080/23311983.2016.1171458>
- Gutiérrez, J. M., García, E., y Ruiz, I. (2018). Narrativas Mediáticas en la Formación Inicial de Educadores: Análisis de una Propuesta de Educación Mediática a través del Uso de Tecnobiografías. *Revista Interuniversitaria de Formación del Profesorado*, 91(32.1), 69-81.
- Haunschild, R., Leydesdorff, L., Bornmann, L., Hellsten, I., y Marx, W. (2019). Does the public discuss other topics on climate change than researchers? A comparison of explorative networks based on author keywords and hashtags. *Journal of Informetrics*, 13(2), 695-707. <https://doi.org/10.1016/j.joi.2019.03.008>
- Haustein, S. (2016). Grand challenges in altmetrics: Heterogeneity, data quality and dependencies. *Scientometrics*, 108(1), 413-423. <https://doi.org/10.1007/s11192-016-1910-9>
- Jackson, M. O. (2010). *Social and economic networks*. Princeton university press.
- Jacomy, M., Venturini, T., Heymann, S., y Bastian, M. (2014). ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software. *PLOS ONE*, 9(6), e98679. <https://doi.org/10.1371/journal.pone.0098679>
- Krotov, V., y Silva, L. (2018). Legality and Ethics of Web Scraping. *AMCIS 2018 Proceedings*. <https://aisel.aisnet.org/amcis2018/DataScience/Presentations/17>
- Ladyman, J., Lambert, J., y Wiesner, K. (2013). What is a complex system? *European Journal for*

- Philosophy of Science*, 3(1), 33-67. <https://doi.org/10.1007/s13194-012-0056-8>
- Landherr, A., Friedl, B., y Heidemann, J. (2010). A Critical Review of Centrality Measures in Social Networks. *Business & Information Systems Engineering*, 2(6), 371-385. <https://doi.org/10.1007/s12599-010-0127-3>
- Ledford, H. (2020). How Facebook, Twitter and other data troves are revolutionizing social science. *Nature*, 582(7812), 328-330. <https://doi.org/10.1038/d41586-020-01747-1>
- Littman, J., Chudnov, D., Kerchner, D., Peterson, C., Tan, Y., Trent, R., Vij, R., y Wrubel, L. (2018). API-based social media collecting as a form of web archiving. *International Journal on Digital Libraries*, 19(1), 21-38. <https://doi.org/10.1007/s00799-016-0201-7>
- Liu, H., Morstatter, F., Tang, J., y Zafarani, R. (2016). The good, the bad, and the ugly: Uncovering novel research opportunities in social media mining. *International Journal of Data Science and Analytics*, 1(3), 137-143. <https://doi.org/10.1007/s41060-016-0023-0>
- Marin, A., y Wellman, B. (2014). Social Network Analysis: An Introduction. En *The SAGE Handbook of Social Network Analysis* (pp. 11-25). SAGE Publications Ltd. <https://doi.org/10.4135/9781446294413>
- Mayr, P., y Weller, K. (2016). Think Before You Collect: Setting Up a Data Collection Approach for Social Media Studies. En L. Sloan y A. Quan-Haase (Eds.), *The SAGE Handbook of Social Media Research Methods* (pp. 107-124). SAGE Publications Ltd. <https://doi.org/10.4135/9781473983847.n8>
- McCay-Peet, L., y Quan-Haase, A. (2017). What is social media and what questions can social media research help us answer? En *The SAGE Handbook of Social Media Research Methods* (pp. 13-26). Sage Thousand Oaks, CA. <https://dx.doi.org/10.4135/9781473983847.n2>
- Menashy, F., y Verger, A. (2019). Network analysis, education policy, and international development: An introduction. *International Journal of Educational Development*, 64, 58-61.
- Myers, S. A., Sharma, A., Gupta, P., y Lin, J. (2014). Information network or social network? The structure of the twitter follow graph. *Proceedings of the 23rd International Conference on World Wide Web*, 493-498. <https://doi.org/10.1145/2567948.2576939>
- Rabbany, R., Takaffoli, M., y Zaiane, O. R. (2011). Analyzing participation of students in online courses using social network analysis techniques. *Proceedings of Educational Data Mining*, 21-30.
- Riquelme, F., y González-Cantergiani, P. (2016). Measuring user influence on Twitter: A survey. *Information Processing & Management*, 52(5), 949-975. <https://doi.org/10.1016/j.ipm.2016.04.003>
- Robinson-Garcia, N., Arroyo-Machado, W., y Torres-Salinas, D. (2019). Mapping social media attention in Microbiology: Identifying main topics and actors. *FEMS Microbiology Letters*, 366(fnz075). <https://doi.org/10.1093/femsle/fnz075>
- Robinson-Garcia, N., van Leeuwen, T. N., y Ràfols, I. (2018). Using altmetrics for contextualised mapping of societal impact: From hits to networks. *Science and Public Policy*, 45(6), 815-826. <https://doi.org/10.1093/scipol/scy024>
- Rodríguez, J. A. (1995). *Análisis estructural y de redes. Colección Cuadernos metodológicos, Núm. 16*. Centro de investigaciones sociológicas.
- Roesslein, J. (2020). *tweepy: Twitter library for python* (3.8.0) [Biblioteca de Python]. <http://github.com/tweepy/tweepy>
- Schroeder, R. (2014). Big Data and the brave new world of social media research. *Big Data & Society*, 1(2), 2053951714563194. <https://doi.org/10.1177/2053951714563194>
- Schuster, J., Jörgens, H., y Kolleck, N. (2019). The rise of global policy networks in education: Analyzing Twitter debates on inclusive education using social network analysis. *Journal of Education Policy*, 1-22. <https://doi.org/10.1080/02680939.2019.1664768>

- Sebei, H., Hadj Taieb, M. A., y Ben Aouicha, M. (2018). Review of social media analytics process and Big Data pipeline. *Social Network Analysis and Mining*, 8(1), 30. <https://doi.org/10.1007/s13278-018-0507-0>
- Shields, R. (2016). Following the leader? Network models of “world-class” universities on Twitter. *Higher Education*, 71(2), 253-268. <https://doi.org/10.1007/s10734-015-9900-z>
- Sloan, L., y Quan-Haase, A. (Eds.). (2017). *The SAGE Handbook of Social Media Research Methods*. SAGE Publications Ltd. <http://dx.doi.org/10.4135/9781473983847>
- Snee, H. (2008). Web 2.0 as a social science research tool. *British Library*, 4(November), 1-34.
- Stieglitz, S., Mirbabaie, M., Ross, B., y Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*, 39, 156-168. <https://doi.org/10.1016/j.ijinfomgt.2017.12.002>
- Thomson, S. D. (2016). *Preserving Social Media* (DPC Technology Watch Report). Digital Preservation Coalition. <http://dx.doi.org/10.7207/twr16-01>
- Traag, V. A., Waltman, L., y van Eck, N. J. (2019). From Louvain to Leiden: Guaranteeing well-connected communities. *Scientific Reports*, 9(1), 5233. <https://doi.org/10.1038/s41598-019-41695-z>
- West, D. B. (2000). *Introduction to Graph Theory*. Pearson.
- Wickham, H., François, R., Henry, L., Müller, K., y RStudio. (2020). *dplyr: A Grammar of Data Manipulation* (1.0.0) [Paquete de R]. <https://CRAN.R-project.org/package=dplyr>
- Zhuhadar, L., y Butterfield, J. (2014). *Analyzing students logs in open online courses using SNA techniques*.