
Representación de las redes sociales

PID_00279380

Jordi Morales-i-Gras

Tiempo mínimo de dedicación recomendado: 3 horas



**Jordi Morales-i-Gras**

Doctor en Sociología por la Universidad del País Vasco (UPV/EHU). Profesor de Análisis de redes, *machine learning* y *big data*, y socio director de Network Oversight, empresa especializada en el análisis sociológico de datos masivos.

El encargo y la creación de este recurso de aprendizaje UOC han sido coordinados por las profesoras: Sílvia Martínez-Martínez, Lola Costa-Gálvez

Primera edición: febrero 2021
© de esta edición, Fundació Universitat Oberta de Catalunya (FUOC)
Av. Tibidabo, 39-43, 08035 Barcelona
Autoría: Jordi Morales-i-Gras
Producción: FUOC
Todos los derechos reservados



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia Creative Commons de tipo Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0. Se puede copiar, distribuir y transmitir la obra públicamente siempre que se cite el autor y la fuente (Fundació per a la Universitat Oberta de Catalunya), no se haga un uso comercial y ni obra derivada de la misma. La licencia completa se puede consultar en: <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

Introducción.....	5
Objetivos.....	6
1. Adquisición de datos de los medios sociales para el análisis de redes.....	7
1.1. Redes sociales y API	7
1.2. API no oficiales y raspado web	13
2. Casos de análisis de redes de los medios sociales.....	16
2.1. Redes de un solo modo	17
2.1.1. Redes de actores	17
2.1.2. Redes de conceptos	23
2.2. Redes de dos modos	27
2.2.1. Redes bipartitas	27
2.2.2. Transformación de las redes de dos modos en redes de un modo	30
3. Limitaciones del análisis de redes sociales, aspectos éticos y de justicia de datos.....	32
Bibliografía.....	35

Introducción

En el módulo «Comprensión de las redes sociales», nos hemos familiarizado con el concepto de *red social* y con el análisis de redes sociales (ARS). Ahora ya sabemos que se trata de un grupo extenso de técnicas que nos permiten analizar y entender conjuntos de datos relacionales. También sabemos que los datos relacionales son especialmente habituales en internet y en los medios sociales, de modo que una gran parte de las interacciones que establecen las personas usuarias con las plataformas digitales, con sus contenidos y con el resto de las usuarias se pueden conceptualizar a modo de red o grafo.

En este módulo, veremos de qué maneras podemos conseguir los datos para llevar a cabo diferentes tipos de análisis de redes, basándonos en los distintos tipos de contenidos que hay disponibles en los medios sociales. Examinaremos cómo conseguir los datos directamente de sus fuentes oficiales o desde fuentes alternativas, y consideraremos cuáles son las implicaciones legales y de legitimidad de todas estas técnicas. Más adelante, mostraremos una serie de casos de análisis, tanto de redes de un modo como de redes bipartitas, y veremos cómo podemos transitar de una perspectiva a la otra. Finalmente, pondremos especial atención en los aspectos éticos que debemos tener en cuenta cuando analizamos las redes, tanto en términos generales como en el caso específico de los medios sociales.

Datos relacionales

Redes de relaciones o vínculos entre varios actores sociales u otros tipos de nodos.

Objetivos

El análisis de redes sociales consiste en un grupo extenso de técnicas que nos permiten analizar y entender conjuntos de datos relacionales. Con el estudio de este módulo deberán alcanzarse los objetivos siguientes:

- 1.** Conocer qué maneras tenemos de conseguir los datos para llevar a cabo diferentes tipos de análisis de redes basándonos en los distintos tipos de contenidos que hay disponibles en los medios sociales.
- 2.** Explorar las diversas estrategias para sintetizarlos a partir de distintos formatos de información.
- 3.** Aprender a generar conocimiento en entornos digitales mediante algunas de las técnicas más beneficiosas y prometedoras de los últimos años y que nos permitirán responder a preguntas relevantes en cualquier escenario de datos.

1. Adquisición de datos de los medios sociales para el análisis de redes

Una gran parte de la información que es capaz de generar y almacenar un ordenador está estructurada a modo de red. No hace tantos años que se usaba la expresión *red de redes* para referirnos a internet, que, al fin y al cabo, es un proyecto construido a base de redes de computadoras conectadas por medio de protocolos TCP/IP. En realidad, si nos fijamos con suficiente atención, veremos que las redes están por todas partes y que hay varias maneras de analizarlas y de generar valor a partir de estas.

El ARS ha experimentado un auténtico renacimiento a raíz de la invención y la expansión de internet, y más particularmente, de los datos masivos y los medios sociales, a causa de que una gran parte de la información que estos contienen y disponen se puede analizar con este tipo de técnicas. En este apartado, veremos cómo podemos obtener datos de los medios sociales más populares de internet, tanto a partir de los servicios oficiales como a partir de fuentes alternativas, y también veremos cómo representarlos y analizarlos mediante herramientas de ARS.

1.1. Redes sociales y API

La mayoría de las plataformas de medios sociales, como por ejemplo Facebook, Instagram, LinkedIn, Twitter o YouTube, disponen de un tipo de servicio denominado *interfaz de programación de la aplicación* o API,¹ que nos sirve para poder establecer una conexión con la aplicación y para intercambiar con esta cierta información.

⁽¹⁾Del inglés, *application programming interface*.

Las API son conjuntos de rutinas que simplifican el intercambio de información entre una aplicación y una persona analista externa o una programadora, y que, a la vez, permiten que la plataforma decida qué parte del total de los datos están a disposición de terceros y con qué condiciones.

Como veremos, no todas las plataformas ofrecen el mismo tipo de datos, ni tampoco han tomado las mismas decisiones respecto a sus API durante los últimos años.

Las API de las plataformas de medios sociales suelen contar con varias **capas de valor**, gratuitas y de pago, a partir de las cuales se pueden obtener datos de la plataforma o, si se trata de una aplicación con permisos de escritura, introducirlos. Los principales usuarios y usuarias de las API son los **analistas de**

datos y los programadores de aplicaciones. Debido sobre todo a los segundos, muchas plataformas permiten, además de descargar datos, publicar contenidos, establecer relaciones e interactuar con contenidos de otros usuarios y usuarias por medio de su API oficial. Estas API, por lo tanto, son un **elemento central de la propuesta de valor** de las propias plataformas. La API es un elemento fundamental a partir del cual las diferentes empresas y productos que utilizan datos de las plataformas establecen conexión, lo que permite crear valor para el usuario final con servicios de analítica, servicios publicitarios y otros servicios específicos.

Las API oficiales de las plataformas de medios sociales son, en teoría, las fuentes de información más estables a las que un analista puede acceder. Sin embargo, lo cierto es que durante los últimos años ha habido una serie de cambios bastante importantes en muchas de estas, que han dejado de proporcionar ciertos datos que antes sí que proporcionaban para adaptarse a las normativas relativamente recientes, como el Reglamento General de Protección de Datos de la Unión Europea (RGPD). En general, la tendencia de los últimos años ha sido la de mostrar un mayor respeto y consideración por la **privacidad de los usuarios y usuarias**, y también, como consecuencia del hecho anterior, la de limitar o dificultar la generación de conocimiento para los investigadores no corporativos.

Cambios en las API oficiales de las plataformas de medios sociales

En el caso de **Twitter**, la API 1.1 no ha sufrido grandes modificaciones, desde que fue lanzada, en cuanto a la descarga de datos, pero sí que ha habido un endurecimiento en cuestiones como, por ejemplo, el registro de nuevas aplicaciones (Hughes, 2018). En cambio, en el caso de **Facebook** (Pérez, 2018), **Instagram** (Gea, 2018) y **LinkedIn** (Wright, 2015), los cambios han sido mucho más profundos, y han dejado de proporcionar varios datos de publicaciones, seguidores y compromiso, restringiendo muchísimo el acceso a las aplicaciones de terceros y la posibilidad de descargar datos. Como resultado de estos cambios, la diferencia entre lo que se podía hacer hace unos años y lo que se puede hacer hoy en estas plataformas es muy grande. Además, no se puede descartar que haya más cambios en un futuro que todavía limiten más el margen de acción de los analistas externos a las plataformas.

Dejando de lado cuestiones como la publicación de mensajes y la interacción con las plataformas y, por lo tanto, centrándonos en la función de las API oficiales de proveedores de datos, típicamente podremos obtener datos con las características siguientes:

1) Perfiles de usuario: información sobre el propietario o propietaria de una cuenta. Cada plataforma dispone de diferentes tipos de información, más o menos completa, y esto también afecta al tipo de datos que están a disposición de terceros. Las redes que disponen de más información sensible, como por ejemplo Facebook o LinkedIn, son las que tienen API más restrictivas. Las que disponen de menos información, como Twitter, son las más abiertas.

Referencias bibliográficas

- M. Hughes** (2018). «Twitter to Place New Restrictions on its API to Stop Abuse» [en línea]. *The Next Web*. Ámsterdam.
- S. Pérez** (2018). «Facebook Rolls out More API Restrictions and Shutdowns» [en línea]. *TechCrunch*.
- L. Gea** (2018). «Instagram restringe el acceso a aplicaciones desarrolladas por terceros» [en línea]. *Octoly*.
- M. Wright** (2015). «LinkedIn Takes Aim at Developers with Plans to Lock Down Most of its APIs» [en línea]. *The Next Web*. Ámsterdam.

2) **Conexiones:** información sobre los contactos de un usuario o usuaria. Hoy en día, solo Twitter ofrece esta información por medio de su API oficial, a pesar de que se trata de datos bastante penalizados en sus tiempos de adquisición: según se estipula en su documentación, la API de Twitter permite la descarga de una serie de conexiones con pausas de quince minutos cuando se logra un cierto volumen de datos.

3) **Publicaciones:** contenidos publicados por los usuarios y usuarias. Es un dato disponible en la mayoría de las API, a pesar de que de distintas maneras y con varios tipos de restricciones. En Facebook, se pueden obtener las publicaciones hechas por páginas y por usuarios en grupos, pero no las publicaciones hechas por los usuarios en sus muros. En Instagram y en LinkedIn, solo se pueden obtener las propias publicaciones con la API estándar. Una vez más, Twitter es la plataforma que actualmente mantiene una política de acceso más abierta, que permite acceder a las últimas 3.200 publicaciones de cada usuario de la plataforma, salvo que este haya declinado la explotación de datos por parte de terceros en su configuración personal.

4) **Interacciones:** información sobre el tipo de vínculos que establecen los usuarios y usuarias entre sí y con los contenidos de otros usuarios. Esta información varía mucho de una red a otra. En Instagram y en LinkedIn, solo es posible acceder a los propios «me gusta»; en Facebook, se pueden obtener las cifras de compromiso y de comparticiones para los apuntes de las páginas públicas y los «me gusta» emitidos por una página, y en Twitter se puede obtener información sobre los retuits y *favs* o «me gusta» a partir de diferentes protocolos internos. Twitter también proporciona información sobre las menciones mantenidas entre los usuarios, los *hashtags* utilizados y sobre las respuestas y citaciones a otros tuits, dado que se trata de datos especialmente interesantes para el ARS.

5) **Grupos y listas:** datos sobre distintas agrupaciones que se pueden llevar a cabo en las redes sociales. En Facebook, es posible acceder a datos sobre grupos específicos, aunque no a datos particulares sobre sus usuarios y usuarias. En Twitter, se puede acceder a las listas creadas por los usuarios y también se puede saber en cuántas de estas ha sido incluido un usuario o usuaria particular.

6) **KPI y datos publicitarios:** datos sobre el rendimiento de las campañas publicitarias y otras cuestiones, como por ejemplo el crecimiento en seguidores de un usuario o usuaria. La mayoría de las plataformas tienen este tipo de datos bastante abiertos en cuanto al rendimiento de la propia cuenta, a la vez que son absolutamente opacos cuando se trata de mostrar los datos de otros usuarios. Esto permite a sus anunciantes poder hacer un seguimiento de las campañas desde fuera de la plataforma y hacer los cruces oportunos con conjuntos diversos, al mismo tiempo que les garantiza que ningún otro usuario o usuaria pueda acceder a este tipo de datos tan sensibles.

Tiempo de descarga

Como mínimo, las relaciones en una población de unas cinco mil personas será obtenida después de una semana entera de descarga ininterrumpida, a pesar de que la cifra variará en función del número de conexiones totales que haya en la población. Para mayor información, podéis ver «Follow, search, and get users», Twitter.

Datos de otros usuarios

Para hacer un seguimiento del crecimiento de las cuentas de otros usuarios o usuarias hay que utilizar servicios específicos y de pago que son ajenos a las plataformas de medios sociales o, si se dispone de datos históricos, elaborar los datos mediante los metadatos presentes en publicaciones capturadas en tiempo real a lo largo de los meses o años.

7) **Metadatos:** datos asociados a los demás conjuntos de datos y que incorporan aspectos como por ejemplo la tecnología utilizada, la resolución de las fotografías y vídeos, el número de seguidores del autor o autora en el momento de la publicación o, en algunos casos, datos sobre aspectos de la personalización de la plataforma llevada a cabo por el usuario: colores, lengua, etc. Estos datos están presentes en todas las API y son los más voluminosos, a pesar de que presentan una gran variabilidad y, a menudo, no aportan grandes elementos de análisis más allá de cuestiones muy técnicas.

Como hemos podido ver, las API oficiales nos pueden proporcionar grandes cantidades de datos muy variados si nos conectamos a ellas.

A la hora de establecer conexión con una API oficial es importante leer bien la documentación relativa a los desarrolladores de aplicaciones de cada plataforma.

Generalmente, hará falta que nos demos de alta como desarrolladores y que creamos una aplicación para acceder a la API. Se trata de un proceso que puede ser más largo en algunas plataformas que en otras, en función de si la plataforma tiene que validar manualmente o no la aplicación que hemos creado. A partir de aquí, será necesario que generemos varios tipos de claves –claves del propietario de las aplicaciones y claves del usuario de las aplicaciones– que deberemos utilizar para establecer la conexión.

El proceso de establecer una conexión con una API puede resultar un poco complicado, sobre todo la primera vez, y por eso hay una serie de servicios que nos ayudarán a **simplificar la conexión**; son los siguientes:

1) Los lenguajes de programación como R, Python o Node disponen de **librerías específicas** para acceder a prácticamente cualquier plataforma de internet, que facilitan la interacción con la API si se dispone de ciertos conocimientos de código. Esta es la alternativa que proporciona una mayor versatilidad y más posibilidades de personalización del servicio, pero también es la opción más complicada de implementar y la que requiere un mayor esfuerzo de mantenimiento. A pesar de que el software que utilizaremos será gratuito, a la hora de implementar una solución de este tipo es importante considerar los costes de un servidor que, idealmente, estará capturando datos de manera permanente.

2) También hay servicios pensados **para analistas sin conocimientos de programación**, pero son bastante más inestables y limitados en sus funciones y capacidades. Algunos de estos son los complementos de Excel Next Analytics o NodeXL, o la aplicación web TAGS de Martin Hawksey. Estas aplicaciones suelen ser gratuitas o tener costes muy bajos, y permiten descargar los datos para llevar a cabo análisis, a pesar de que no suelen tener cuadros de mando

API públicas vía Python

Algunos usuarios de GitHub mantienen el siguiente listado de API públicas accesibles vía Python.

de análisis automático ni visualizaciones demasiado atractivas. Otro problema habitual es que no son herramientas pensadas para una captura de datos permanente, sino más bien para usos esporádicos.

3) Finalmente, también es importante conocer el mundo de las **plataformas de social listening**, en las que compiten empresas como Netbase, Brandwatch, Salesforce, IBM Watson, Talkwalker, Websays, Atribus o Brand24, entre muchas otras. Son servicios muy sencillos de utilizar, dirigidos a equipos de comunicación y marketing, que incorporan varias capas de valor a los datos mediante herramientas de procesamiento del lenguaje natural, como por ejemplo el análisis de sentimiento o mediante algoritmos de aprendizaje automático. En general, se trata de servicios muy estables, que capturan datos de manera permanente y que permiten su análisis prácticamente en tiempo real. Se trata de servicios de pago, muchos de ellos con un coste bastante o muy elevado. No obstante, la mayoría de estos servicios no facilitan la descarga de datos y deberemos limitarnos al análisis incrustado en la aplicación web.

En función del tipo de necesidades de análisis que tengamos y de factores como el tipo de datos que necesitemos, su frecuencia o su volumen, tendremos que optar por uno u otro servicio. Si elegimos el primer tipo de servicio, deberemos complementarlos con un software de transformación y carga de datos (es decir, ETL o ELT), como por ejemplo Talend o Apache NiFi, que nos permita incrustar el código generado y ejecutarlo de manera recurrente, y también transformar los datos y almacenarlos de la manera deseada. En este sentido, es importante que tengamos en cuenta que los datos que provienen de los medios sociales presenten un **formato semiestructurado** cuando son extraídos de la API y, por lo tanto, debemos proceder a estructurarlos para introducirlos en una base de datos relacional como MySQL o PostgreSQL (ved figura 1) o en una base de datos no relacional como MongoDB o Neo4j.

En cualquier caso, y sea cual sea el sistema de acceso a la API que elegimos llevar a cabo, siempre es muy importante que tengamos en cuenta los aspectos relativos al uso de los datos que prevén las propias plataformas y que imponen a cualquiera de los usuarios y usuarias de las API mediante el **contrato de acceso** y el **documento de condiciones de uso**.

Por regla general, en ningún caso podremos llevar a cabo análisis que faciliten la identificación de individuos en términos ideológicos, religiosos, médicos o psicológicos. Según la mayoría de los contratos, tampoco podremos llevar a cabo tareas de espionaje, vigilancia o seguridad. Todas estas normas no son simples compromisos éticos –que, por otro lado, todo analista debería tener–,

sino que se trata de contratos que el analista establece con las plataformas bajo la base legal del RGPD y cuyo incumplimiento puede comportar efectos jurídicos.

Las **API oficiales** son la fuente de información más estable en términos técnicos a la que un o una analista de datos de los medios sociales puede recurrir. Se trata de servicios robustos que proporcionan grandes cantidades de datos que pueden inyectarse con posterioridad en distintos programas de ETL o ELT, para llevar a cabo el tipo de análisis que sea necesario, siempre que este análisis entre dentro de lo que se prevé en las condiciones de uso de las propias API y en el RGPD. La principal desventaja de las API oficiales es la inestabilidad permanente que deriva no de la arquitectura del servicio, sino del hecho de que la plataforma pueda decidir dejar de proporcionar ciertos datos en cualquier momento y hacerlo de manera más o menos arbitraria, transparente y justificada.

Figura 1. Ejemplo de estructuración de los datos de Twitter

JSON (algunos campos)

```
{
  "created_at": "Thu Oct 01 00:00:00 +0000 2020",
  "id_str": "803166678308339418",
  "text": "Ejemplo de tuit que menciona @UOCuniversitat y @UOCinfocom",
  "user": {
    "id": 444810854,
    "name": "Jordi Morales i Gras",
    "screen_name": "jmoralesigras",
    "location": "Barcelona"
  },
  "entities": {
    "user_mentions": [
      {
        "name": "UOCuniversitat",
        "screen_name": "UOCuniversitat",
        "id": 14919552
      },
      {
        "name": "UOC Estudis InfoCom",
        "screen_name": "UOCinfocom",
        "id": 1283569494
      }
    ]
  }
}
```

Tabla de tuits

Id del tuit	Autor	Fecha (día)	Texto	Ubicación
803166678308339418	jmoralesigras	01-10-2020	Ejemplo de tuit que menciona @UOCuniversitat y @UOCinfocom	Barcelona
...	...	...	...	...

Tabla de menciones

Id del tuit	Autor	Mención
803166678308339418	jmoralesigras	UOCuniversitat
803166678308339418	jmoralesigras	UOCinfocom
...	...	...

Fuente: elaboración propia con datos de ejemplo

1.2. API no oficiales y raspado web

A causa de los cambios en las políticas de acceso a los datos de las API oficiales de las plataformas de medios sociales, y como estrategia alternativa para la obtención de datos estructurados, durante los últimos años han proliferado los servicios basados en **API no oficiales** y en técnicas como el **raspado web** (*web scraping*). El concepto de «no oficialidad» no equivale al de ilegalidad o ilegitimidad, sino que se trata simplemente de servicios que no están mantenidos oficialmente por una plataforma. De hecho, como veremos más adelante, muchas API no oficiales requerirán que nos identifiquemos en la plataforma de medios sociales con nuestro nombre de usuario y clave y, por lo tanto, se dedicarán a explotar una funcionalidad no explotada por la API oficial con el beneplácito de la propia plataforma o, como mínimo, dejando constancia de ello y sin que la plataforma lo impida.

Las API no oficiales que proporcionan datos sobre los medios sociales son un tipo muy particular que permiten descargar datos o llevar a cabo otras operaciones, como transformarlos en un *feed* RSS, previsualizar y descryptar enlaces acortados, transformar los formatos de las imágenes y vídeos o, incluso, detectar bots y otros tipos de usuarios con comportamientos automatizados.

Muchas de estas API se pueden adquirir en *marketplaces* específicos y otras se pueden encontrar en plataformas para compartir código fuente como GitHub.

A la hora de acceder a los contenidos de los medios sociales, algunas API no oficiales necesitarán que les proporcionemos nuestra identificación de usuarios o usuarias. Algunas otras, en cambio, no lo necesitarán porque se limitarán a acceder a datos públicamente disponibles para internautas no identificados. La diferencia entre estos dos tipos de API no oficiales es clave.

- Por un lado, las API no oficiales que **requieren un usuario identificado** están sujetas a las normas de uso que impone la plataforma. En estos casos, y como usuario identificado, el o la analista tendrá que acatar las limitaciones de volumen o frecuencia, entre otras condiciones que variarán según cada plataforma.
- Por otro lado, en el caso de las API que devuelven datos públicamente disponibles y que **no requieren un usuario identificado**, el o la analista tendrá algo más de libertad para acceder sin restricciones a la totalidad de los datos que la plataforma haya dispuesto para un usuario no identificado. En estos casos, y siempre que no se haya aceptado un texto legal, el analista no estará sujeto a la normativa específica de la plataforma.

Servicios de API no oficiales

En el mundo de las API no oficiales, hay servicios de todo tipo que proporcionan datos muy diversos, como por ejemplo datos climáticos, deportivos, de movilidad, culturales, financieros o de medios digitales, entre otros muchos casos. También son habituales las API de este tipo que ofrecen servicios como la traducción de textos o la implementación de algoritmos clasificatorios sobre los datos.

Enlaces complementarios

Algunos de los *marketplaces* son: RapidAPI o PROMPT API. Algunos usuarios de GitHub mantienen el listado unofficial-apis de API no oficiales.

Pero no estar sujetos a la normativa de una plataforma no implica tener carta blanca para llevar a cabo los análisis que se quieran. En cualquier caso, el o la analista seguirá estando sujeto a las leyes de propiedad intelectual y al RGPD, que están por encima de las normativas específicas de las plataformas. Esto implica que las técnicas de raspado web se tendrán que aplicar sobre las bases legitimadoras del interés público o del interés legítimo y que el analista tendrá que ser proactivo en la toma de decisiones técnicas y metodológicas que favorezcan el cumplimiento de la normativa, como por ejemplo la anonimización irreversible o la agregación de casos.

Las API no oficiales que acceden a los datos públicos para usuarios no identificados lo hacen mediante una técnica llamada *raspado web* (*web scraping*). En esencia, la técnica consiste en automatizar –y, normalmente, ocultar detrás de un servidor intermediario– las visitas a un sitio web y copiar y pegar sus contenidos públicos en una base de datos.

Se trata de una técnica ampliamente utilizada por todo tipo de servicios de internet, desde indexadores de contenidos, como Google o Bing, hasta comparadores de productos de todo tipo: perfumes, zapatos, seguros, viajes, hoteles, etc. También se trata de técnicas utilizadas en el marco del denominado *hacking ético* y que permiten, por ejemplo, indexar los contenidos ilegales que se encuentran en la web profunda y ponerlos a disposición de las autoridades.

Pero no todas las actividades de raspado web son legales o legítimas. Es con esta misma técnica con la que se llevan a cabo una gran variedad de actividades malintencionadas, como por ejemplo la adquisición de correos electrónicos con el fin de enviar correos basura y software malicioso, o con el fin de captar datos personales que permitan hacer cruzados con otras bases de datos. Utilizar una API no oficial basada en raspado web o crear un robot que lleve a cabo una tarea de este tipo es una alternativa que el analista podrá considerar para la adquisición de datos, pero deberá tener en cuenta todos los aspectos anteriores sobre la **legitimidad del análisis** y, en caso de duda, consultarlo con un servicio jurídico especializado.

Con las API no oficiales y el raspado web, a menudo, es posible acceder a datos que las API oficiales no proporcionan pero que son, al fin y al cabo, datos públicos y accesibles para cualquier usuario no identificado. Sin embargo, hay que tener en cuenta que se trata de datos que se pueden ver afectados por aspectos como la configuración regional del navegador, y también que siempre serán datos menos ricos en metadatos que los datos de las API oficiales.

Los datos de Twitter

En el caso de los datos de Twitter, con las API no oficiales y el raspado web no podremos obtener los nombres de los usuarios y usuarias que hayan retuiteado un tuit –algo que sí que proporciona la API oficial–, sino que nos tendremos que conformar con conocer la cifra total. En cambio, podremos llevar a cabo análisis retroactivos sin demasiadas

Lecturas recomendadas

Sobre la legalidad del raspado web en España, podéis ver:

B. Martín Fernández (2019, 17 de junio). «El ‘web scraping’ y la protección de datos» [en línea]. *Cinco Días*.

Sobre el principio de responsabilidad proactiva de los responsables del tratamiento de datos personales, podéis ver:

J. A. Vega Ortega (2019, 1 de octubre). «Por qué es tan importante en la pyme el principio de responsabilidad proactiva» [en línea]. *Cinco Días*.

complicaciones, mientras que la API oficial nos limitará las búsquedas a los últimos siete o nueve días.

Las API no oficiales y la técnica del raspado web son fuentes de datos más inestables que las fuentes oficiales y, por regla general, menos ricas en cuanto al tipo de datos que proporcionan. Además, como se trata de fuentes no oficiales, la responsabilidad legal de la adquisición y el tratamiento de los datos recae totalmente en el analista, puesto que la plataforma no es responsable, más allá de haber dispuesto la información públicamente, algo que, en principio, habrá hecho asesorada por sus servicios jurídicos. Como cuestiones positivas, hay que destacar que este tipo de técnicas proporcionan más autonomía al analista y lo hacen depender algo menos de las decisiones de las plataformas en materia de política de datos, a pesar de que, al fin y al cabo, el raspado web anónimo dejará de ser posible en el momento en el que las plataformas decidan que ningún usuario no identificado pueda visualizar ningún tipo de contenido.

2. Casos de análisis de redes de los medios sociales

Tal como vimos en el módulo «Comprensión de las redes sociales», el **ARS** consiste en un grupo de técnicas de análisis que permiten generar conocimiento y valor sobre conjuntos de datos relacionales. Se trata de técnicas especialmente pertinentes para analizar los datos que recogemos de los medios sociales y que tienden a representar las relaciones que se establecen entre los usuarios y usuarias (por ejemplo, menciones, relaciones de amistad o de seguimiento, etc.) o que se establecen entre los usuarios y los contenidos (por ejemplo, publicaciones, páginas web, productos, *hashtags*, etc.).

Durante los últimos años, han proliferado una gran variedad de estudios e investigaciones llevadas a cabo con datos de las API oficiales de los medios sociales² y mediante técnicas de ARS (Ediger y otros, 2010; Chatfield y Brajawidagda, 2012):

⁽²⁾Generalmente datos de Twitter, por el hecho de que es la red más abierta a la investigación académica.

- Por un lado, destacan lo que podemos denominar *redes de actores*, en las que se estudian relaciones de distintos tipos establecidas entre usuarios de los medios sociales. Dentro de esta tipología, destacan las **redes de menciones entre usuarios** (Tremayne, 2014; Congosto, 2015; Del Fresno y otros, 2015; Morales-i-Gras, 2015; Apodaka y Morales-i-Gras, 2016) y los análisis sobre las **relaciones de amistad o seguimiento** (Myers y otros, 2014; Grandjean, 2016).
- Por otro lado, destacan lo que podemos denominar *redes de conceptos*, en las que se exploran **relaciones entre palabras, *hashtags* u otros tipos de elementos** propios del discurso en los medios sociales (Zhao, 2013; Gualda y Borrero, 2015; Himelboim y otros, 2017; Gerlach, Peixoto y Altman, 2018).

El ARS es un grupo de técnicas muy extenso y que permite trabajar con una gran variedad de datos de los medios sociales, tanto provenientes de las API oficiales como de API no oficiales o del raspado web.

En función del tipo de datos de los que disponemos, y del tipo de técnicas de transformación de datos que articulamos, podremos llevar a cabo unos análisis u otros. Seguidamente, veremos y comentaremos unos cuantos ejemplos.

2.1. Redes de un solo modo

Las **redes de un solo modo** son aquellas en las que todos sus nodos representan un mismo tipo de entidad, a pesar de que hayan podido sintetizarse a partir de redes de dos modos.

Seguidamente veremos unos cuantos ejemplos de redes de actores en las que todos los nodos representan a usuarios y usuarias de la misma plataforma, y también de redes de conceptos, en las que todos los nodos representan palabras o *hashtags* utilizados en la conversación mantenida en una plataforma de medios sociales. Para analizar las redes, nos basaremos en varios conceptos que vimos en el módulo «Comprensión de las redes sociales», como la diferencia entre las redes dirigidas y no dirigidas o las redes ponderadas y no ponderadas.

2.1.1. Redes de actores

Las redes que se detallan en este apartado tienen como protagonistas a los **usuarios y usuarias de las redes sociales**, a quienes denominaremos *actores*. Las redes de actores pueden ser según Lozares (1996):

- **Redes egocéntricas**, en las que se parte de un actor central y se visualiza el entorno.
- **Redes sociocéntricas**, que no parten de ningún actor particular, sino de un sistema de relaciones.

Veamos algunos ejemplos, entre los cuales hay algunos que forman parte de la historia de internet porque ya no es posible llevarlos a cabo con las API actuales.

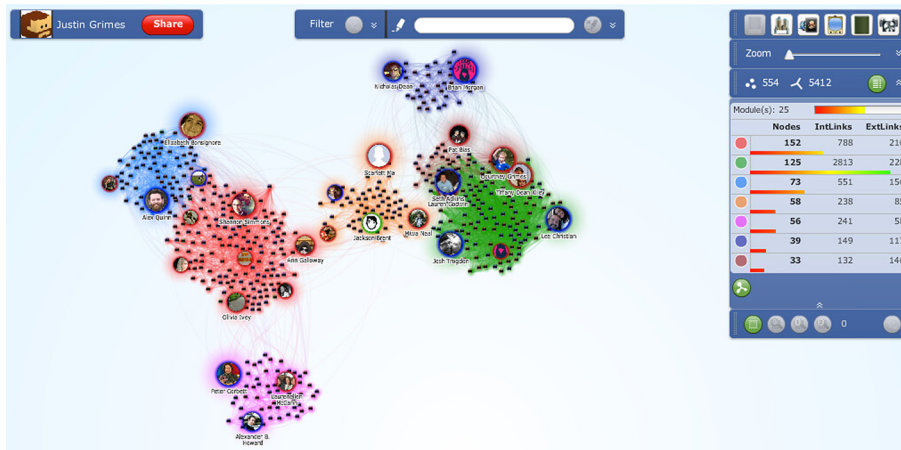
Red de amistades en Facebook

En la red, están representadas las amistades entre los amigos de un usuario de Facebook. Se trata de una red de tipo **egocéntrica, no dirigida y no ponderada**, elaborada a partir de las relaciones de amistad en Facebook con la aplicación Challenger (figura 2). Facebook ya no proporciona los datos necesarios para sintetizar este tipo de grafo desde su API oficial y la aplicación Challenger ya no existe. Los nodos de la red han sido coloreados según un algoritmo de detección comunitaria y han sido ponderados según el grado de entrada.

Referencia bibliográfica

C. Lozares Colina (1996).
«La teoría de redes sociales».
Papers: revista de sociologia
(vol. 48, págs. 103-126).

Figura 2. Red de amistades en Facebook

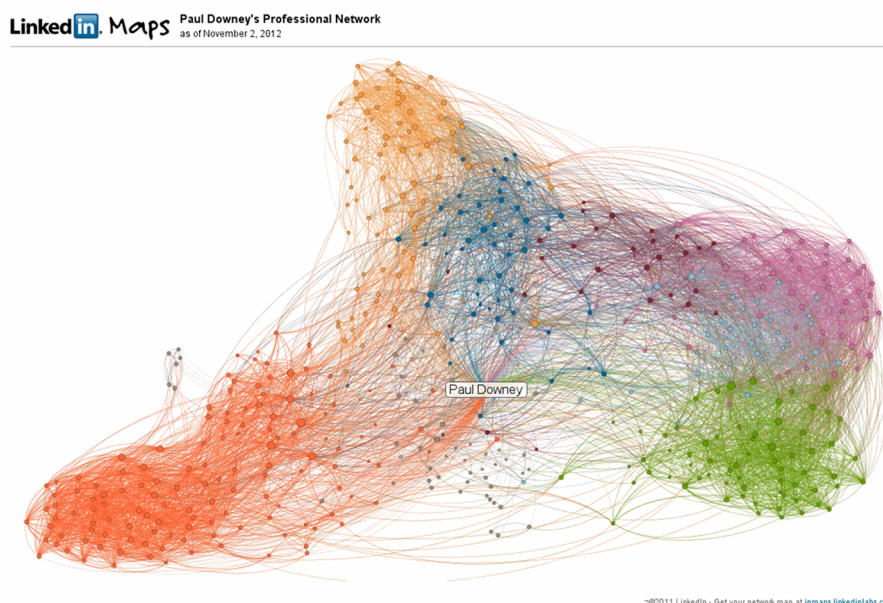


Fuente: flickr

Red de amistades en LinkedIn

Como en el caso anterior, se trata de una red **egocéntrica**, **no dirigida** y **no ponderada**, elaborada a partir de las relaciones de amistad en LinkedIn entre los amigos de un usuario con la aplicación LinkedIn Maps (figura 3). Como Facebook, LinkedIn ya no proporciona los datos necesarios para llevar a cabo este tipo de análisis y LinkedIn Maps ya forma parte de la historia de internet.

Figura 3. Red de amistades en LinkedIn

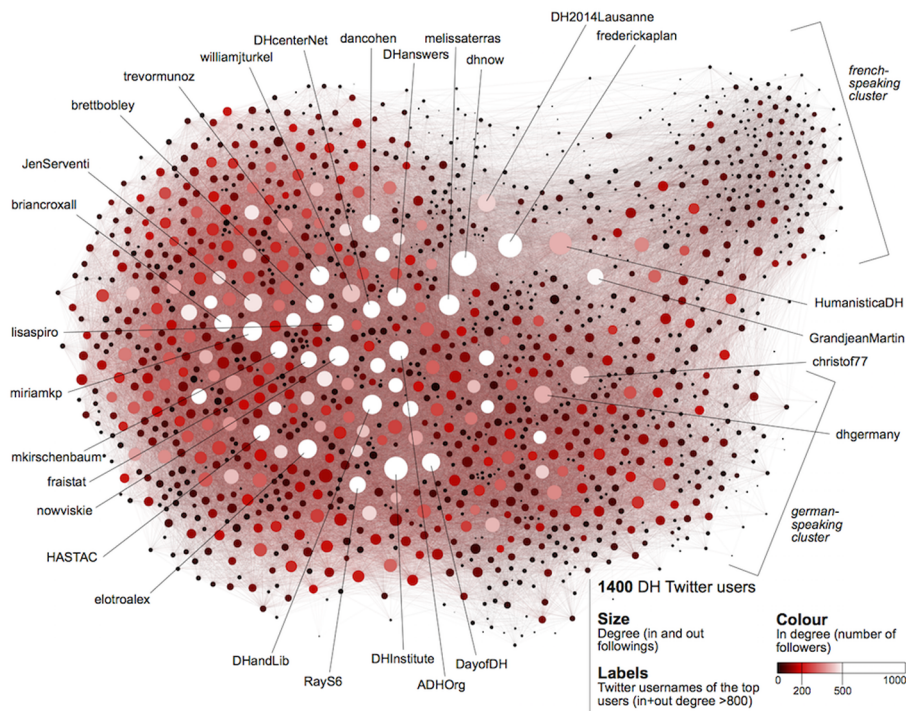


Fuente: flickr

Red de relaciones en Twitter entre los seguidores de un usuario

En la red siguiente, están representadas las relaciones de seguimiento entre los seguidores de un usuario de Twitter (figura 4). Se trata de una red **egocéntrica**, **no ponderada** y **dirigida**, a diferencia de las anteriores, porque en Twitter las relaciones de seguimiento no es necesario que sean recíprocas como en Facebook o LinkedIn. La red ha sido sintetizada con el software Gephi y los nodos

Figura 5. Red de relaciones en Twitter: miembros de una comunidad



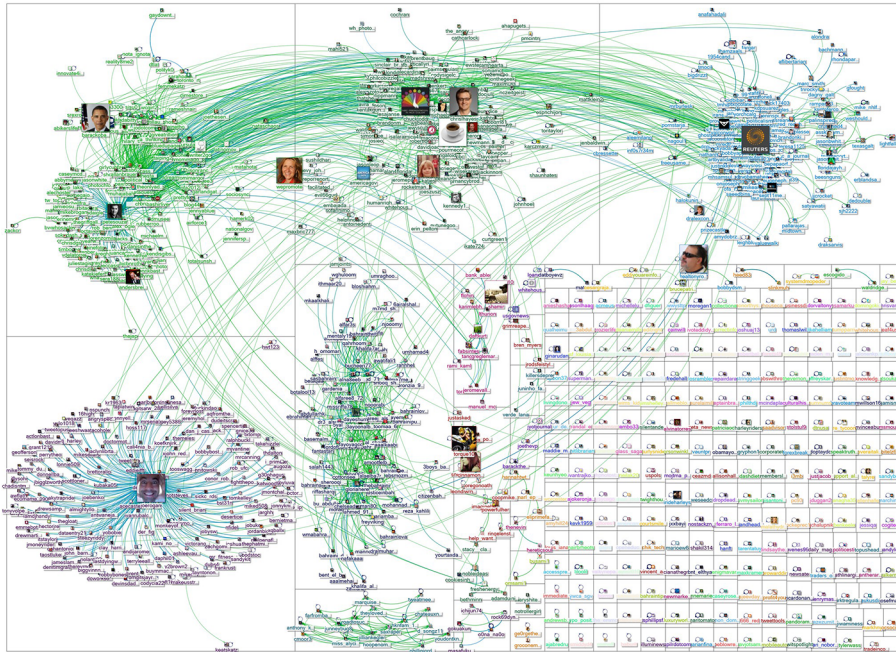
Fuente: Wikimedia commons

Red de relaciones en Twitter entre participantes con un mismo hashtag

Esta red también es **sociocéntrica y dirigida**, y recoge las relaciones de seguimiento y las respuestas entre los usuarios que usaron el *hashtag* #whitehouse en Twitter el 30 de diciembre de 2011 (figura 6). Dado que recoge más de un tipo de interacción posible y las respuestas son un tipo de interacción que puede darse varias veces, se trata de una red **ponderada**. Este tipo de grafos sirven para observar el fenómeno conocido como *cámaras de eco* (DiFonzo, 2011), según el cual los usuarios y usuarias acaban participando de aquellas conversaciones en las que participa su círculo de amistades y gente sociológicamente parecida, lo que contribuye a reforzar la propia opinión y tal vez a confundirla con la opinión dominante en una sociedad. La red ha sido creada con el software NodeXL y el algoritmo Clauset-Newman-Moore (2004).

Referencia bibliográfica

A. Clauset; M. E. Newman; C. Moore (2004). «Finding Community Structure in Very Large Networks». *Physical review E* (vol. 6, núm. 70, 066111).

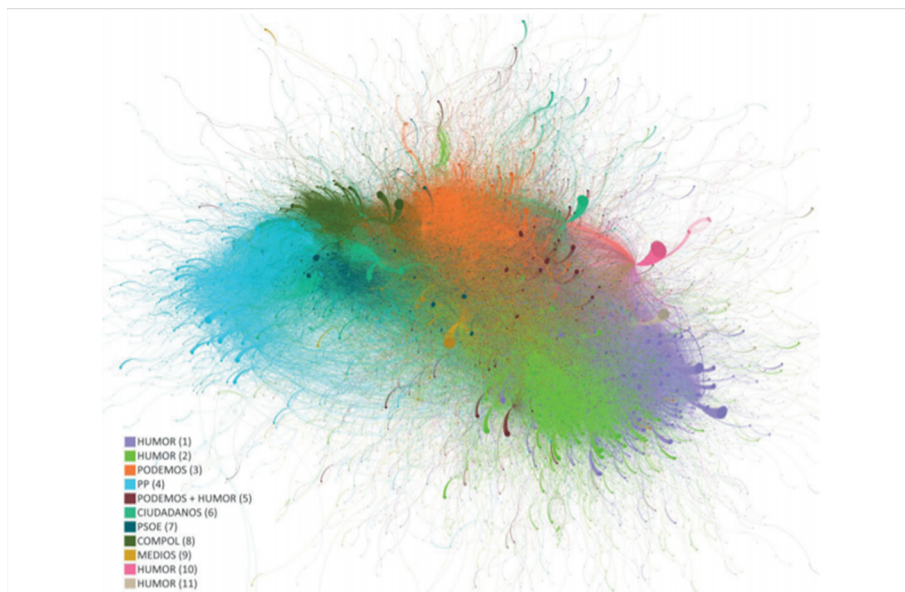
Figura 6. Red de relaciones en Twitter: participantes con un mismo *hashtag*

Fuente: flickr

Red de menciones en Twitter durante un debate político televisado

La red masiva de la figura 7 representa 273.238 menciones entre 92.052 usuarios y usuarias que participaron en un debate electoral televisado en las elecciones generales españolas el 28 de abril de 2019. La red es **dirigida y ponderada**, puesto que algunos usuarios han mencionado a otros muchas veces. Los usuarios han sido coloreados según la comunidad identificada con el algoritmo Louvain y representan a comunidades de usuarios que se han agrupado alrededor de varios candidatos o de perfiles humorísticos que hacían bromas y memes sobre el debate. El grafo ha sido elaborado con Gephi.

Figura 7. Red de menciones en Twitter en un debate político televisado

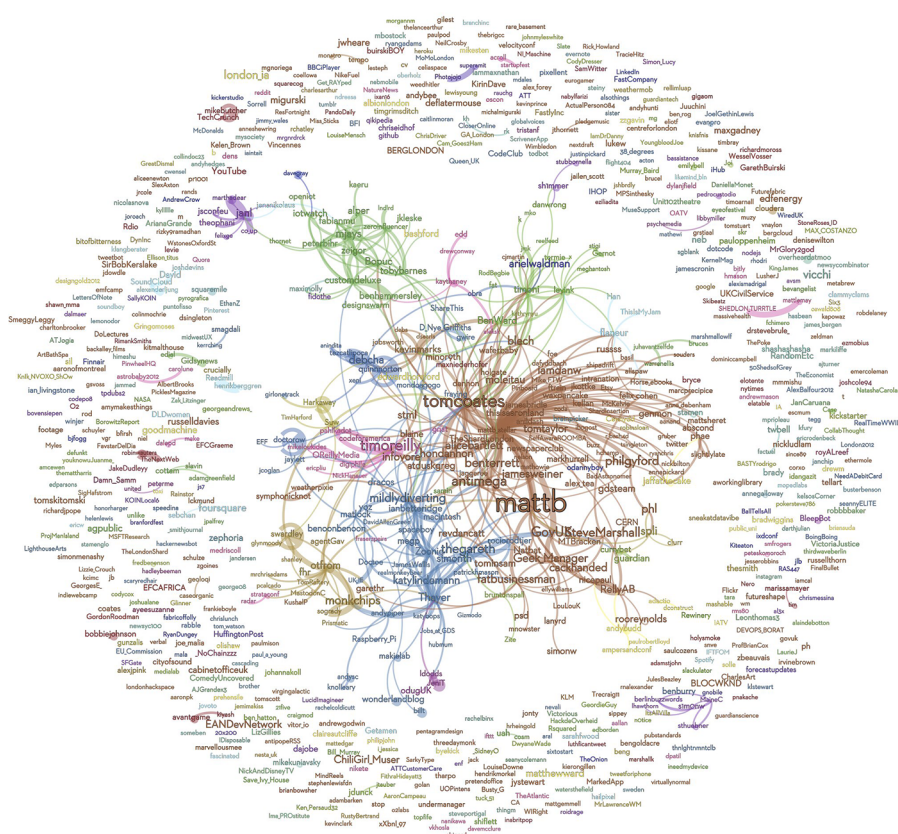


Fuente: J. Orbegozo-Terradillos; A. Larrondo-Ureta; J. Morales-i-Gras (2020). «Influencia del género en los debates electorales en España: análisis de la audiencia social en #ElDebateDecisivo y #L6Neldebate». *El profesional de la información (EPI)* (vol. 29, núm. 2).

Red de menciones en Twitter entre usuarios seguidos por un usuario

Esta red refleja las menciones mantenidas en Twitter entre usuarios y usuarias seguidos por un usuario o usuaria (figura 8). A diferencia del grafo anterior, aquí no se trata de ver las comunidades que emergen en un debate masivo, sino de identificar hasta qué punto un usuario sigue a usuarios que tienden a participar de los mismos debates y a retuitarse y responderse mutuamente. El grafo, **dirigido y ponderado**, ha sido elaborado con Gephi y las comunidades han sido identificadas con el algoritmo Louvain.

Figura 8. Red de menciones en Twitter entre usuarios seguidos por un usuario



Fuente: flickr

2.1.2. Redes de conceptos

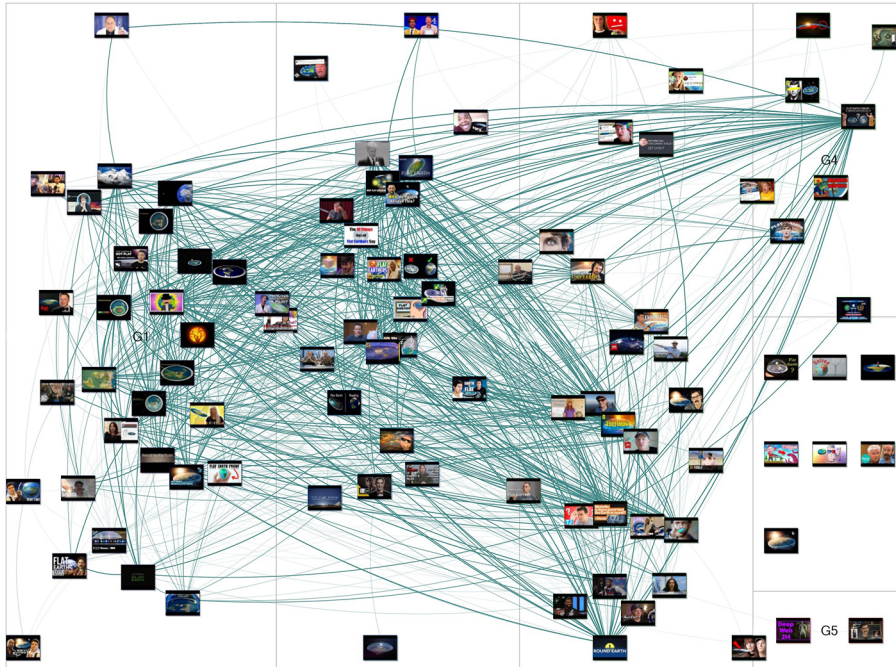
Los nodos de las redes siguientes no representan a actores, sino conceptos que han sido utilizados por los usuarios en varias conversaciones en los medios sociales: **palabras** y **hashtags**. Con este tipo de aproximaciones, el ARS también se presenta como una herramienta para el análisis del discurso y el modelado temático: los algoritmos de aprendizaje no supervisado orientados a la identificación de temas en un texto o conjunto de textos.

Red de conceptos en Twitter

En esta red, podemos ver tanto los *hashtags* como las palabras que han aparecido conjuntamente en los tuits publicados con el *hashtag* #dataviz, durante la primera semana de octubre de 2020 (figura 9). Se trata de una red de tipo **ego-céntrica** en cuanto al *hashtag* #dataviz. Es una red **no dirigida**, puesto que la coocurrencia entre palabras es necesariamente recíproca y **ponderada**, porque las combinaciones son más o menos comunes en la conversación. Los colores de las palabras han sido determinados con el algoritmo Louvain y el grafo ha sido sintetizado con Gephi.

NodeXL y el algoritmo Clauset-Newman-Moore (2004). Este análisis se puede llevar a cabo hoy en día con datos de la API de YouTube, que son accesibles con la versión de pago del complemento de Excel NodeXL.

Figura 12. Red de vídeos en Youtube



Fuente: flickr

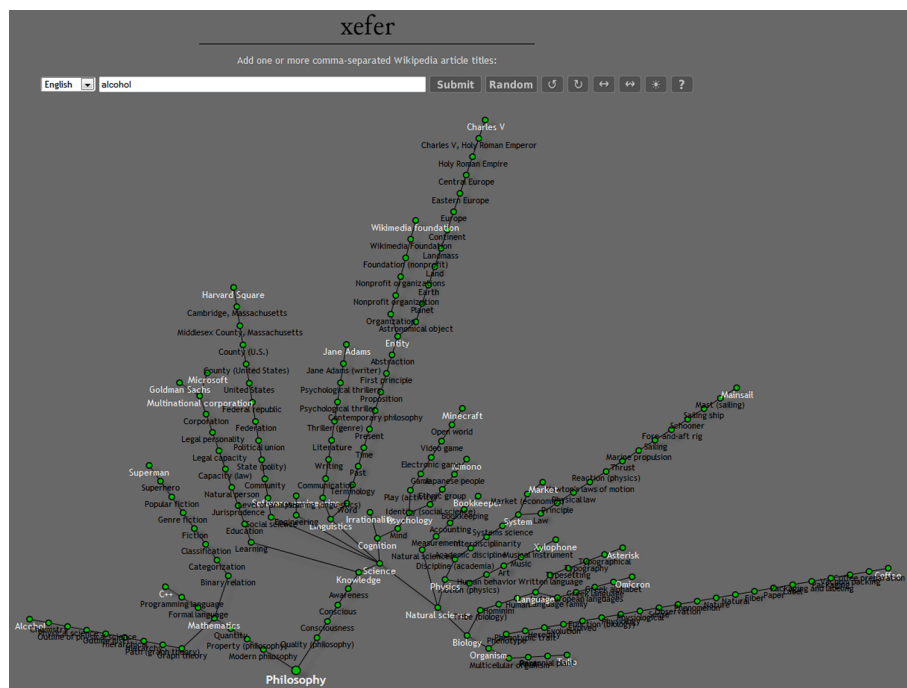
Red de artículos de Wikipedia

La red que muestra la figura 13 ha sido elaborada con una curiosa aplicación web gratuita que demuestra que el 97 % de los artículos de Wikipedia nos conducen al artículo denominado *Filosofía* si hacemos clic de manera recursiva en el primer enlace del artículo. El fenómeno se conoce como *getting to philosophy* y lo puede experimentar cualquier internauta por medio de la web del proyecto Xefer. En términos técnicos, se trata de un grafo radial **no dirigido** y **no ponderado**, a pesar de que su interpretación es la de un grafo dirigido en el que, empecemos por la palabra que empecemos, acabaremos llegando a la filosofía.

Enlace complementario

Podéis consultar «Getting to Philosophy» en la propia Wikipedia.

Figura 13. Red de artículos de Wikipedia que apuntan al de *Filosofía*



Fuente: Wikimedia commons

2.2. Redes de dos modos

Todas las redes anteriores, ya sean redes de actores o conceptuales, tienen como protagonistas nodos de una sola entidad. En este apartado, veremos redes que comprenden más de un tipo de nodos. En la literatura de ARS, es habitual denominar estas redes como *redes bipartitas*, o *redes de actores y eventos* (Reichardt y otros, 2011). Presentaremos una serie de ejemplos, y también veremos cómo las podemos transformar en redes de un modo, que son, al fin y al cabo, el tipo de redes más sencillas de interpretar y de las que disponemos de un mayor volumen de métricas y algoritmos.

2.2.1. Redes bipartitas

Las redes bipartitas representan las relaciones establecidas entre nodos que representan diferentes tipos de objetos.

Como veremos, este tipo de representación nos permite convertir prácticamente cualquier base de datos en una red sobre la que podremos aplicar, con posterioridad, una serie de transformaciones y técnicas de análisis.

Red de editores de imágenes en Wikipedia

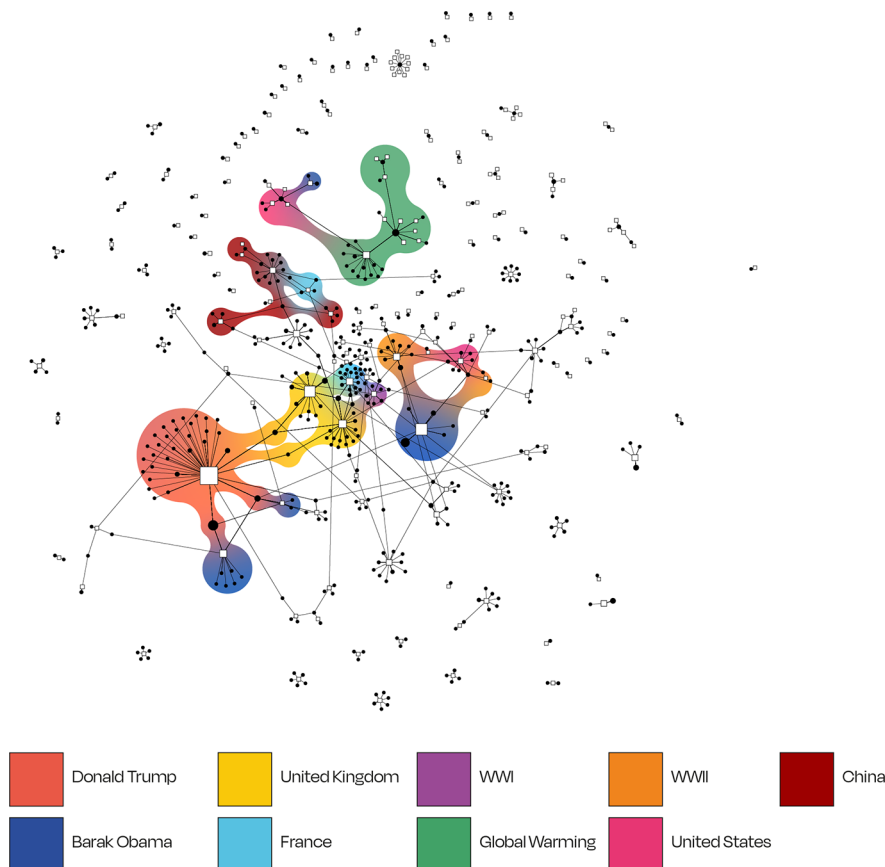
La red que muestra la figura 14 recoge las imágenes con visualizaciones de datos de las cien páginas más vistas de Wikipedia en inglés, representadas con cuadrados, y de sus editores, representados con círculos. Las medidas de los cuadrados han sido **ponderadas** según el número de veces que se ha actuali-

Referencia bibliográfica

J. Reichardt; R. Alaminio; D. Saad (2011). «The Interplay between Microscopic and Mesoscopic Structures in Complex Networks» [en línea]. *PLoS One* (núm. 6, e21282).

zado una visualización de datos, y las medidas de los círculos representan el número de veces que un usuario ha editado una imagen. Los colores del grafo se utilizan a modo de leyenda para indicar a qué artículos pertenecen las imágenes que han sido modificadas más veces y por más editores, lo que nos da una idea de cuáles son los temas más polémicos.

Figura 14. Red de editores de imágenes en Wikipedia



Fuente: Wikimedia commons

Red de usuarios y *hashtags* en Instagram

La red bipartita representada en la figura 15 recoge una serie de usuarios de Instagram relativos a empresas españolas y una serie de *hashtags* vinculados con la pandemia de la COVID-19 que estas empresas utilizaron entre los meses de febrero y marzo de 2020. Se trata de una red **no dirigida y ponderada**, en la que se han coloreado los nodos en función de si se trata de empresas o de *hashtags*. Los datos para llevar a cabo este análisis se han obtenido mediante técnicas de raspado web con el servicio Apify.

2.2.2. Transformación de las redes de dos modos en redes de un modo

Ya hemos visto unos cuantos ejemplos de cómo el ARS nos permite investigar y generar conocimiento según las relaciones establecidas entre usuarios, entre contenidos o entre usuarios y contenidos de los medios sociales. Hemos visto que hay redes de dos modos, en las que se tienen en cuenta relaciones entre entidades diferentes (por ejemplo, una red de usuarios y *hashtags*) y que las herramientas estándares de ARS nos ayudan a poderlas visualizar de varias maneras y enfatizar diferentes tipos de propiedades y características.

La mayoría de las métricas y algoritmos de ARS están pensados y funcionan mejor cuando se aplican sobre redes de un solo modo. Por ello, es habitual llevar a cabo **procesos de transformación** de las redes de dos modos en redes de un modo.

Tal y como comentamos en el módulo «Comprensión de las redes sociales», se trata de procedimientos que pueden resultar matemáticamente complejos y, por esa razón, el software específico de ARS cuenta con herramientas que permiten transformar las redes de dos modos o bipartitas en redes de un solo modo.

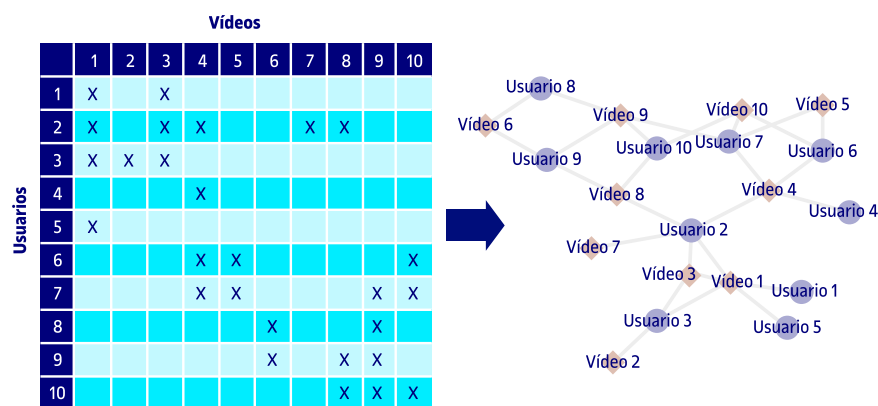
Procesos de transformación

Por ejemplo, si disponemos de un conjunto de datos sobre un listado de usuarios que han visto una serie de vídeos, podremos interpretar este conjunto de datos como una matriz de doble entrada y, en consecuencia, como una red de dos modos. En la red de la figura 17, que ha sido creada con Pajek, distinguimos usuarios y vídeos con nodos con forma de círculo o de rombo. A partir de esta red de dos modos, se nos presentan dos escenarios diferentes si queremos transformar la red en un solo modo:

- 1) Sintetizar una red de usuarios en función de los vídeos que han visto.
- 2) Sintetizar una red de vídeos en función de los usuarios que los han visto.

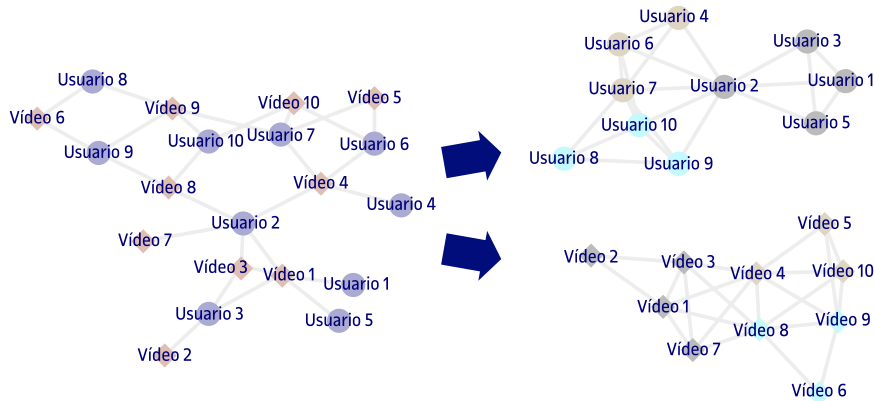
Ambas opciones se pueden implementar (ved la figura 18) y nos permiten desplegar las técnicas de ARS optimizadas para redes de un solo modo.

Figura 17. Diseño de una red de dos modos a partir de una matriz de doble entrada



Fuente: elaboración propia con datos aleatorios y Pajek

Figura 18. Diseño de una red de dos modos a partir de una matriz de doble entrada



Fuente: elaboración propia con datos aleatorios y Pajek

Mediante esta técnica de transformación de datos, el abanico de opciones para la aplicación de técnicas de ARS, según los datos de los medios sociales y de internet en su conjunto, se multiplica exponencialmente. Siempre que dispongamos de un conjunto de datos que se pueda disponer a modo de filas y columnas, será posible establecer una relación entre los casos contenidos en las filas y las variables dispuestas en las columnas a modo de red de dos modos o bipartita. Este tipo de transformación, aplicado de manera correcta, nos puede resultar muy útil a la hora de transformar los datos para cumplir con el RGPD y la normativa de propiedad intelectual –lo que dará lugar a agregaciones irreversibles de datos que permiten estudiar fenómenos sociales sin poner en peligro los datos personales o sensibles de los usuarios–, incluso en el supuesto de que los datos sean adquiridos mediante fuentes no oficiales.

3. Limitaciones del análisis de redes sociales, aspectos éticos y de justicia de datos

Siempre que analizamos datos de tipo social hemos de tener en cuenta que detrás hay personas y, por lo tanto, la **sostenibilidad ética** de cualquier investigación es fundamental y es un aspecto que siempre hay que tener en cuenta: en la fase de diseño de una investigación, durante la recogida de datos, durante el análisis y durante la fase de aplicación de los resultados, si se da el caso.

Recordad: es importante reflexionar y entender los **diversos aspectos éticos** que implica una investigación con datos sociales –y los datos de los medios sociales son datos sociales– tanto en cuanto a la obtención de los datos como a su tratamiento y su análisis.

Según dictamina el Informe Belmont (US Department of Health and Human Services, 1979), las investigaciones con personas han de ser respetuosas y deben garantizar su autonomía, informándolas **por escrito y bajo firma** de las condiciones, del alcance y de los objetivos de la investigación, permitiéndolas dejar de formar parte de esta en cualquier momento y sin consecuencias negativas de ningún tipo. Estas investigaciones también se han de regir según el **principio de beneficencia**, que implica beneficiar o, como mínimo, no perjudicar a los y las participantes de un estudio. Por último, los procedimientos tienen que ser **justos, razonables e inteligibles** para los y las participantes en los estudios. Obviamente, se trata de principios que, a menudo, son muy difíciles de satisfacer mediante el análisis de redes sociales, incluso en un entorno fuera de línea.

Para empezar, en una **red social de tipo egocéntrica**, sí que será posible conseguir el **consentimiento informado** de una persona (esto es, *ego*), pero será prácticamente imposible conseguir el consentimiento de todos los demás nodos de la red, que serán nombrados por *ego*, y cuyas relaciones quedarán a cuerpo descubierto. Otro problema habitual es que a estos individuos nombrados por *ego* no se les garantiza el derecho a no participar en el estudio, puesto que, de hecho, participan sin ni siquiera saberlo. Tradicionalmente, los analistas de las redes personales han resuelto estos temas problemáticos mediante pseudónimos o anonimizando las redes y agregando datos siempre que ha sido posible. Se trata de soluciones claramente subóptimas y con unos resultados muy discutibles en cuanto al cumplimiento de los principios del Informe Belmont.

El Informe Belmont

El Informe Belmont recopila una serie de principios éticos para la investigación con seres humanos, creado para las ciencias biomédicas, pero que a menudo se reconoce como referente científico general.

US Department of Health and Human Services (1979).
The Belmont Report [en línea].

Es importante entender, por lo tanto, que los problemas de orden ético con el análisis de redes sociales son muy anteriores a los datos masivos y a los medios sociales. De hecho, con los datos provenientes de los medios sociales tenemos una clara ventaja, como mínimo en términos legales, en el hecho de que los usuarios y usuarias que forman parte de una plataforma de medios sociales –a cuyos datos podemos acceder de una manera u otra–, todos y todas **han firmado unos términos de servicio** con la plataforma en cuestión, en los que se especifican los usos analíticos por parte de terceros. Esto implica que los datos dispuestos en una API oficial son datos de usuarios que han autorizado su explotación por parte de terceros, siempre que se cumplan una serie de condiciones, como por ejemplo la no elaboración de algoritmos que identifiquen las categorías sensibles. El caso del raspado web es muy parecido, siempre que se practique desde un usuario no identificado, puesto que los datos que una plataforma dispone para cualquier individuo, en principio, tienen que ser datos de usuarios que hayan autorizado este tipo de uso. De no ser así, sería la plataforma la que incumpliría el RGPD.

Pero, a pesar de que los términos de uso de las plataformas garanticen la legalidad de ciertos análisis, desde un punto de vista de la ética investigadora, es importante tener en cuenta el resto de los principios que van más allá de los consentimientos informados. Las investigaciones con datos sociales pueden tener una gran capacidad de **impacto social**, y por lo tanto es importante velar por que este impacto sea positivo y beneficioso para los objetos de estudio y para la sociedad en general. También es importante que los procesos de investigación sean justos y que guarden un buen equilibrio de costes y beneficios, al tiempo que deben ser inteligibles para la sociedad en su conjunto. Por eso, también es muy importante tener en cuenta la **función divulgadora** de la investigación y el **principio de retorno** a la sociedad.

Bibliografía

Apodaka, E.; Morales-i-Gras, J. (2016). «Redes solidarias en Twitter: un acercamiento a la estructura del independentismo catalán en base a datos capturados en Twitter». *Virtualis* (vol. 14, núm. 7, págs. 53-89).

Chatfield, A. T.; Brajawidagda, U. (2012). «Twitter Tsunami Early Warning Network: a Social Network Analysis of Twitter Information Flows». *Proceedings of the 23rd ACIS Conference* (págs. 1-10).

Clauset, A.; Newman, M. E.; Moore, C. (2004). «Finding Community Structure in Very Large Networks». *Physical review E* (vol. 6, núm. 70, 066111).

Congosto, M. L. (2015). «Elecciones europeas 2014: viralidad de los mensajes en Twitter». *Redes. Revista hispana para el análisis de redes sociales* (vol. 1, núm. 26, págs. 23-52).

Del-Fresno-García, M.; Daly, A. J.; Supovitz, J. (2015). «Desvelando climas de opinión por medio del social media mining y análisis de redes sociales en Twitter. El caso de los Common Core State Standards». *Redes. Revista hispana para el análisis de redes sociales* (vol. 1, núm. 26, págs. 53-75).

DiFonzo, N. (2011). «The echo-chamber effect». *New York Times* (núm. 22).

Ediger, D.; Jiang, K.; Riedy, J.; Bader, D. A.; Corley, C.; Farber, R.; Reynolds, W. N. (2010). «Massive Social Network Analysis: Mining Twitter for Social Good». *2010 39th International Conference on Parallel Processing* (págs. 583-593).

Gea, L. (2018). «Instagram restringe el acceso a aplicaciones desarrolladas por terceros» [en línea]. *Octoly*. [Fecha de consulta: 18 de noviembre de 2020].

Gerlach, M.; Peixoto, T. P.; Altmann, E. G. (2018). «A Network Approach to Topic Models». *Science Advances* (vol. 7, núm. 4, eaq1360).

Grandjean, M. (2016). «A Social Network Analysis of Twitter: Mapping the Digital Humanities Community». *Cogent Arts & Humanities* (vol. 1, núm. 3, 1171458).

Gualda, E.; Borrero, J. D. (2015). «La “Spanish Revolution” en Twitter (2): redes de hashtags (#) y actores individuales y colectivos respecto a los desahucios en España». *Redes. Revista hispana para el análisis de redes sociales* (vol. 1, núm. 26, págs. 1-22).

Himmelboim, I.; Smith, M. A.; Rainie, L.; Shneiderman, B.; Espina, C. (2017). «Classifying Twitter Topic-Networks Using Social Network Analysis». *Social Media+ Society* (vol. 1, núm. 3).

Hughes, M. (2018). «Twitter to Place New Restrictions on its API to Stop Abuse» [en línea]. *The Next Web*. Ámsterdam. [Fecha de consulta: 18 de noviembre de 2020].

Kralj Novak, P. y otros (2018). «Impact Investing Market on Twitter: Influential Users and Communities» [en línea]. *Applied Network Science* (vol. 10, núm. 3). [Fecha de consulta: 18 de noviembre de 2020].

Lozares Colina, C. (1996). «La teoría de redes sociales». *Papers: revista de sociologia* (vol. 48, págs. 103-126).

Martín Fernández, B. (2019, 17 de junio). «El ‘web scraping’ y la protección de datos» [en línea]. *Cinco Días*. [Fecha de consulta: 18 de noviembre de 2020].

Morales-i-Gras, J. (2015). «Desenredando las identidades soberanistas vasca y catalana: un Análisis de Redes Sociales de las etiquetas de Twitter #BasquesDecide y #Up4Freedom». *Papeles del CEIC. International Journal on Collective Identity Research* (vol. 2, págs. 1-37).

Myers, S. A.; Sharma, A.; Gupta, P.; 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* (págs. 493-498).

Orbegoza-Terradillos, J.; Larrondo-Ureta, A.; Morales-i-Gras, J. (2020). «Influencia del género en los debates electorales en España: análisis de la audiencia social en #ElDebateDecisivo y #L6Neldebate». *El profesional de la información (EPI)* (vol. 29, núm. 2).

Perez, S. (2018). «Facebook Rolls out More API Restrictions and Shutdowns» [en línea]. *TechCrunch*. [Fecha de consulta: 18 de noviembre de 2020].

Reichardt, J.; Alamino, R.; Saad, D. (2011). «The Interplay between Microscopic and Mesoscopic Structures in Complex Networks» [en línea]. *PloS One* (núm. 6, e21282). [Fecha de consulta: 18 de noviembre de 2020].

Schöps, J. D.; Kogler, S.; Hemetsberger, A. (2020). «(De-) Stabilizing the Digitized Fashion Market on Instagram-Dynamics of Visual Performative Assemblages». *Consumption Markets & Culture* (vol. 2, núm. 23, págs. 195-213).

US Department of Health and Human Services (1979). *The Belmont Report* [en línea]. [Fecha de consulta: 18 de noviembre de 2020].

Vega-Ortega, J. A. (2019, 1 de octubre). «Por qué es tan importante en la pyme el principio de responsabilidad proactiva» [en línea]. *Cinco Días*. [Fecha de consulta: 18 de noviembre de 2020].

Wright, M. (2015). «LinkedIn Takes Aim at Developers with Plans to Lock Down Most of its APIs» [en línea]. *The Next Web*. Ámsterdam. [Fecha de consulta: 18 de noviembre de 2020].

Zarco, C.; Cordon, O. (2020). «Analyzing the Communication in Social Media of the Main Sustainable Brands During COVID-19 Crisis: the Spanish vs. Italian Cases» [en línea]. *Research Square*. [Fecha de consulta: 18 de noviembre de 2020].

Zhao, Y. (2013). «Analysing Twitter Data with Text Mining and Social Network Analysis». *Proceedings of the 11th Australasian Data Mining and Analytics Conference* (pág. 23).