
Comprensión de las redes sociales

PID_00279379

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 sociodirector 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. Análisis de redes sociales.....	7
1.1. Conceptualización del análisis de redes sociales	7
1.2. Breve historia del análisis de redes sociales	8
2. Los elementos del análisis de redes sociales.....	11
2.1. Nodos, aristas y grafos	11
2.2. Direccionalidad y ponderación	12
2.3. Redes de uno y dos modos	13
2.4. El lenguaje de las redes	14
3. Principales métricas en el análisis de redes sociales.....	16
3.1. Métricas de los nodos de una red	16
3.2. Métricas de la red en su conjunto	21
4. Visualización de redes sociales.....	26
4.1. Software de análisis de redes sociales	26
4.2. Algoritmos de visualización de redes sociales	27
Bibliografía.....	33

Introducción

La gran mayoría de los datos que se capturan en internet, y muy significativamente en los medios sociales, no suelen ser los datos estructurados que encontraremos en una hoja de cálculo o en una base de datos relacional, en las que hay una serie de columnas que representan las variables y una serie de filas que representan a los individuos, como podría ser una base de datos de clientes o de proveedores, ni tampoco los datos recogidos en un cuestionario de una encuesta. Los datos que se presentan mayoritariamente en los medios sociales son publicaciones, interacciones y conexiones entre los usuarios y usuarias, y para poder analizarlas, hay que recurrir a técnicas como el **análisis de redes sociales** (de ahora en adelante, ARS).

El ARS es una técnica especialmente pertinente 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.).

El ARS consiste en una serie de técnicas orientadas al análisis de los datos interconexionados que centran su atención en las relaciones o conexiones entre un conjunto de actores sociales.

En este módulo, nos familiarizaremos con los elementos fundamentales del ARS y aprenderemos a utilizarlo con diferentes tipos de datos para la generación de conocimiento. Veremos cuáles son sus elementos principales, presentaremos algunas de sus métricas más interesantes y, finalmente, nos centraremos en aspectos como por ejemplo la visualización de redes con software específico, que nos ayudará a comprender y poder explicar qué es lo que sucede en los medios sociales desde una perspectiva de gran valor.

Objetivos

Con el estudio de este módulo deberán alcanzarse los siguientes objetivos:

1. Conocer las técnicas de análisis de las redes sociales y sus elementos fundamentales.
2. Utilizar las métricas más importantes de análisis de las redes sociales.
3. Visualizar las redes mediante un software específico para comprender y poder explicar qué es lo que sucede en los medios sociales.

1. Análisis de redes sociales

Para llevar a cabo el ARS, es importante conocer el concepto de *red social* y su importancia en diferentes disciplinas científicas. También haremos un breve recorrido por su historia, que nos permitirá conocer el origen de muchos de los algoritmos y de las métricas que usaremos más adelante.

1.1. Conceptualización del análisis de redes sociales

El ARS tiene su origen en la ciencia social y en la matemática discreta. Hoy en día, se aplica y se sigue desarrollando en una gran variedad de campos de estudio: desde la física teórica a la antropología social.

A causa de su transdisciplinariedad, a veces el ARS toma diferentes nombres.

- Desde la psicología y las ciencias de la educación, es habitual referirse a él como *sociometría* y suele aplicarse al estudio de las relaciones entre el alumnado de un grupo clase.
- Por otro lado, desde la física y las ciencias de la computación, es más habitual referirse a la *teoría de grafos*. La **teoría de grafos** es la teoría matemática que estudia el concepto de *grafo* o *red*: estructuras de nodos y aristas que modelan las relaciones entre los objetos.

Todos estos conceptos presentan algunas diferencias entre sí, diferencias que tienen que ver, sobre todo, con la tradición disciplinaria de las distintas ciencias y de su momento histórico, pero nos remiten al mismo tipo de perspectiva estructural de análisis de los datos. Aquí, trataremos indistintamente todas estas perspectivas y nos referiremos a *redes* o *grafos* como sinónimos perfectos.

Redes: herramientas de cálculo matemático y, a la vez, representaciones visuales de conjuntos de datos relacionales.

Lo primero que hay que entender es que el ARS implica un modelo de **tratamiento de los datos** que está presente en todos los campos de la ciencia y que consiste en centrar la atención en las **relaciones** o las **conexiones** que se establecen entre los individuos o las entidades que sean objeto del estudio (Freeman, 2012).

Análisis de redes

Por ejemplo, un astrofísico podrá usar el análisis de redes para analizar las dinámicas entre los planetas, un economista podrá analizar los flujos de las importaciones y las exportaciones entre países, y un bioquímico podrá investigar los vínculos entre las diferentes células que componen un organismo. Lo que todos estos análisis tendrán en común es

Sociometría

La sociometría fue fundada por el psiquiatra Jacob Levy Moreno y por la psicóloga social Helen Hall Jennings (Freeman, 2012). Sus estudios fundacionales se basaban en el análisis de las relaciones establecidas entre los individuos en poblaciones acotadas, como por ejemplo grupos de personas presas o alumnado interno.

Teoría de grafos

El psicólogo matemático Frank Harary es considerado el padre de la versión moderna de esta teoría (Freeman, 2012), que fue clave en el nacimiento de la computación y la inteligencia artificial.

Referencia bibliográfica

L. C. Freeman (2012). *El desarrollo del análisis de redes sociales: un estudio de sociología de la ciencia*. Bloomington, Indiana: Palibrio.

que centrarán su atención en los vínculos entre las entidades estudiadas, y no tanto en sus propiedades.

El **análisis relacional o estructural** propio del ARS es diferente y complementario al análisis atributivo o individualista. Mientras que la perspectiva atributiva o individualista centra la actividad empírica en recopilar casos individuales y caracterizarlos mediante una serie de atributos, la perspectiva relacional se basa en recopilar información sobre los **vínculos entre los casos individuales**. Según las preguntas que nos formulemos, deberemos usar un análisis de uno u otro tipo. En cualquier caso, tal y como podemos ver en el ejemplo de la figura 1, siempre será posible transitar desde una perspectiva atributiva a una relacional y viceversa.

Figura 1. Datos atributivos y datos relacionales

Tabla de doble entrada: tres individuos y sus aficiones (0 = no le gusta, 1 = sí que le gusta)				Matriz de adyacencias: número de aficiones que comparten los 3 individuos			
Casos	Ir a la playa	Ir al cine	Tocar el violín		Joan	Laura	Manel
Joan	0	1	0	Joan		1	2
Laura	1	1	1	Laura	1		2
Manel	1	1	0	Manel	1	2	

Esta matriz es simétrica: tantas
aficiones comparte Laura con Joan
como Joan con Laura

Fuente: elaboración propia

En el paradigma de los datos masivos, es muy habitual disponer de datos que se representan mejor en una matriz de adyacencias, como por ejemplo las interacciones entre usuarios o el número de veces que cada usuario ha publicado un mensaje utilizando un *hashtag* particular. Muchas veces, estas acciones e interacciones serán la mejor manera de generar «propiedades» que podamos asignar a los casos en nuestros análisis. Por todo esto, el ARS es un método de análisis que adquiere protagonismo en los entornos digitales, a pesar de que su historia empieza mucho antes de la creación y la expansión de la computación.

1.2. Breve historia del análisis de redes sociales

A pesar de que los filósofos y científicos sociales siempre han tenido la idea intuitiva de que los «lazos» entre los individuos tienen consecuencias sociales importantes, no fue hasta los años treinta del siglo XX cuando un grupo de científicos, inicialmente liderados por los psicólogos sociales Jacob Levi Moreno y Helen Hall Jennings, empezaron a desarrollar una perspectiva empírica, sistemática y matemáticamente refinada que, con el tiempo, se denominaría ARS (Freeman, 2012). Las primeras investigaciones de ARS son, por lo tanto, muy anteriores a la era digital. Si dejamos de lado las investigaciones matemáticas que conforman la prehistoria del ARS –como por ejemplo el problema de

los siete puentes de Königsberg resuelto por Leonhard Euler en 1736–, es posible distinguir **tres fases** en la historia del desarrollo del ARS y solo en la última podemos considerar que los medios sociales tienen un impacto notable.

El problema de los siete puentes

Se trata de un problema matemático clásico resuelto por Euler (1736), que se considera el nacimiento de la topología y la teoría de grafos. En la ciudad prusiana de Königsberg (es decir, la actual Kaliningrado, que pertenece a Rusia) se pretendía recorrer los siete puentes que unen las cuatro partes de la ciudad en un desfile, pero solo pasando una vez por cada puente. Euler demostró que esto no era posible, interpretando cada parte de la ciudad como un nodo y cada puente como una arista. En su solución, solo sería posible recorrer cada puente una vez si hubiera un número par de puentes.

1) Fase de eclosión. Desde el nacimiento del ARS en los años treinta hasta las postrimerías de los años cincuenta del siglo XX, una serie de investigaciones muy débilmente vinculadas entre sí permitieron desarrollar los **conceptos esenciales del ARS**. Los pioneros fueron Jacob Levy Moreno y Helen Hall Jennings, que llevaron a cabo las primeras investigaciones sistemáticas de ARS y empezaron a utilizar explícitamente conceptos como «red» o «díada», para referirse a la relación entre pares de nodos. También destacan las investigaciones de Elisabeth Bott y John Barnes, antropólogos estudiosos de los lazos familiares y de las sociedades urbanas que dieron nombre al ARS. Por otro lado, fue muy importante la investigación de los psicólogos sociales del MIT Kurt Lewin, Alex Bavelas y Dorwin Cartwright, inventores de conceptos como «clíca» (esto es, grupos de nodos perfectamente conectados) o de métricas como el «grado» (esto es, el número de conexiones o de «vecinos» de cada nodo en el grafo). Y también hay que mencionar a su colaborador matemático Frank Harary, formalizador de la teoría de grafos moderna y figura clave en el desarrollo de las ciencias de la computación y la inteligencia artificial.

2) Fase de formalización. Las décadas de los sesenta, setenta y ochenta del siglo XX implicaron una gran **maduración intelectual** para el ARS. En esta fase, el ARS extendió enormemente su radio de aplicación y se utilizó en investigaciones de disciplinas tan diferentes como las ciencias de la comunicación, la historia o los estudios de capital social. Además de la investigación sustantiva y aplicada, también se profundizó en su dimensión formal, con lo que se crearon nuevas métricas, como la «centralidad de intermediación», propuesta por Linton C. Freeman (1977; 1978) y que veremos más adelante, y técnicas de análisis, como el *blockmodeling* (es decir, una técnica estadística que permite generar redes aleatoriamente para contrastar hipótesis), introducida por los sociólogos Harrison White, Scott Boorman y Ronald Breiger (1976). Gracias a todo esto, el ARS fue incluido en varios planes de estudio de las mejores universidades del mundo, como por ejemplo Harvard, Columbia o Stanford. De aquellos científicos, además de sus contribuciones académicas, hay que destacar su papel como profesores y formadores de analistas de redes.

Referencias bibliográficas

- L. C. Freeman (1977). «A set of Measures of Centrality Based on Betweenness». *Sociometry* (págs. 35-41).
- L. C. Freeman (1978). «Centrality in Social Networks Conceptual Clarification». *Social Networks* (vol. 3, núm. 1, págs. 215-239).
- H. C. White; S. A. Boorman; R. L. Breiger (1976). «Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions». *American Journal of Sociology* (vol. 4, núm. 81, págs. 730-780).

ARS aplicada a diferentes disciplinas

A continuación exponemos algunas de las aplicaciones del ARS en investigaciones de diferentes disciplinas:

- **Comunicación:** el psicólogo social Stanley Milgram (1967) demostró con un experimento de campo que todos los habitantes del planeta estamos conectados por una media de seis contactos intermedios.
- **Historia:** el historiador y politólogo Karl Wolfgang Deutsch (1952) utilizó modelos de redes en sus estudios sobre comunicación y nacionalismo para conceptualizar los flujos de información.
- **Capital social:** el sociólogo Mark Granovetter (1973) demostró que los vínculos débiles que establecemos con personas que no forman parte de nuestro círculo inmediato (por ejemplo, antiguas amistades, antiguos compañeros y compañeras de estudios, etc.) tienen un gran potencial de generación de oportunidades laborales, puesto que son una puerta abierta a información que no es habitual en nuestro círculo.

3) Fase de expansión, en la que todavía estamos. A partir de la década de los noventa y con la irrupción de internet, el ARS experimentó un auténtico **renacimiento**. Inicialmente fue el correo electrónico y después fueron los medios sociales, que simplificaron y abarataron los costes de la recopilación de datos de las redes, lo que propició multitud de análisis estructurales. A la vez, y como consecuencia de estas transformaciones tecnológicas, se puso sobre la mesa la necesidad de algoritmos capaces de computar grandes volúmenes de datos de manera rápida y fiable. De este periodo destacan científicos con una gran orientación computacional, como el físico Mark Newman, creador de la métrica de modularidad (2006) que utilizamos para medir la calidad de las comunidades –también denominados *grupos* o *clústeres*– en un grafo masivo; o como Mathieu Jacomy, ingeniero jefe del software Gephi y coautor de algoritmos de visualización de redes masivas, como Force Atlas 2 (Jacomy y otros, 2014).

Por lo tanto, una cuestión muy importante es entender que el análisis de redes sociales y la teoría de grafos son anteriores al concepto de *redes sociales* de internet, tal como las entendemos hoy en día. De hecho, se trata de herramientas que han servido de inspiración para los principales arquitectos de la computación y de la World Wide Web. Del mismo modo, el concepto de *red social* es muy anterior a Facebook o Twitter. Las redes sociales digitales están hechas a imagen y semejanza de las analógicas. Un grupo de compañeros y compañeras de trabajo, de amistades o una familia es, al fin y al cabo, una red social. La especie humana nos relacionamos entre nosotros y nosotras formando redes de relaciones y el ARS es una perspectiva privilegiada que permite entender estas relaciones y sus consecuencias sociales.

Referencias bibliográficas

- S. Milgram** (1967). «The Small World Problem». *Psychology Today* (vol. 1, núm. 2, págs. 60-67).
- K. W. Deutsch** (1952). «On Communication Models in the Social Sciences». *Public Opinion Quarterly* (vol. 3, núm. 16, págs. 356-380).
- M. S. Granovetter** (1973). «The Strength of Weak Ties». *American Journal of Sociology* (vol. 6, núm. 78, págs. 1360-1380).

Referencias bibliográficas

- M. E. Newman** (2006). «Modularity and Community Structure in Networks». *Proceedings of the National Academy of Sciences* (vol. 23, núm. 103, págs. 8577-8582).
- M. Jacomy; T. Venturini; S. Heymann; M. Bastian** (2014). «ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software». *PloS One* (vol. 6, núm. 9, e98679).

2. Los elementos del análisis de redes sociales

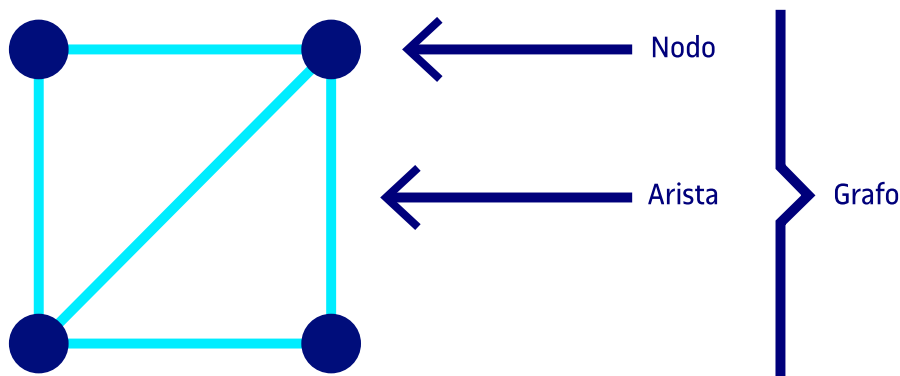
El ARS es una herramienta que nos permite interpretar la complejidad de las relaciones humanas y entender sus consecuencias.

Los análisis de redes siempre se han caracterizado por el **rigor empírico** (Lozares, 1996) y, por lo tanto, lo mejor que podemos hacer es centrarnos inicialmente en sus conceptos clave y dejar la interpretación para más adelante.

2.1. Nodos, aristas y grafos

Cualquier red o grafo está compuesta por dos elementos fundamentales: nodos y aristas (ved figura 2). Los **nodos** de la red –a veces llamados *vértices*, *puntos* o *actores*– son sus entidades. Un nodo puede representar cualquier tipo de entidad (por ejemplo, a personas, empresas, usuarios, webs, etc.). Las **aristas** –también llamadas *enlaces*, *líneas* o *relaciones*– son los vínculos entre estas entidades. Una arista puede representar cualquier tipo de relación (por ejemplo, una mención, una relación de conocimiento, una llamada telefónica, etc.).

Figura 2. Elementos fundamentales de una red o grafo



Fuente: elaboración propia

Un **grafo** se puede utilizar para representar diferentes tipos de relaciones sociales y de esto dependerán sus rasgos fundamentales.

Según las características que tomen las aristas de una red, nos encontraremos ante un tipo u otro de grafo en términos de **direccionalidad** y **ponderación**. En cambio, las características que tomen los nodos determinarán si nos encontramos ante una **red de un solo modo** o de **dos modos**.

Referencia bibliográfica

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

2.2. Direccionalidad y ponderación

Las relaciones que representan las aristas de un grafo pueden tener **dirección** o no tenerla. En función de si las aristas de un grafo representan relaciones dirigidas o no dirigidas, nos encontraremos con redes o grafos dirigidos o no dirigidos.

- **Relaciones no dirigidas.** Una relación podrá considerarse no dirigida cuando sea necesariamente recíproca o bidireccional, o cuando la dirección de la relación no sea considerada importante.

Algunos casos clásicos son la amistad entre dos amigos, el grado de conocimiento entre dos familiares o el número de atributos en común entre dos individuos en un estudio de encuesta. En los medios sociales, tenemos varios casos de relaciones no dirigidas, como por ejemplo las amistades en Facebook o LinkedIn: o hay un vínculo bidireccional o no la hay.

- **Relaciones dirigidas.** Una relación podrá considerarse dirigida cuando no sea necesariamente recíproca, o cuando la dirección sea un factor importante para el análisis.

Casos clásicos son las redes de préstamos monetarios, las llamadas telefónicas o las exportaciones e importaciones. En los medios sociales, son relaciones dirigidas las menciones y los seguimientos en Twitter o los *swipes* en Tinder: yo puedo seguir a alguien que no me sigue o puedo manifestar interés por alguien a quien no intereso.

Ya hemos dicho que la teoría no es lo que más preocupa a los y las analistas de ARS. La mayoría de las veces, el sentido común es el que nos ayudará a conceptualizar una relación como dirigida o no dirigida. Otras veces no lo tendremos tan claro y deberemos recurrir a la literatura científica de referencia para adaptarnos a un estándar concreto. Ahora bien, en cualquiera de los casos, es muy importante entender que es el o la analista quien tiene que tomar la decisión de si una red será interpretada como dirigida o no dirigida (por ejemplo, podemos asumir que dos personas se conocen simétricamente, o que una de las dos conoce más a la otra). La decisión que tome el analista deberá ser congruente con las técnicas analíticas que se apliquen con posterioridad. Por lo tanto, queramos o no, siempre necesitaremos un mínimo de teoría que nos ayude a interpretar una red de relaciones.

Además de la dirección, hay un segundo elemento que es clave a la hora de determinar el tipo de aristas que componen un grafo, se trata del peso o la **ponderación** de las aristas. Las aristas no ponderadas dan lugar a grafos no ponderados, y las aristas ponderadas, a grafos ponderados.

- **Relaciones no ponderadas.** Un conjunto de relaciones podrán considerarse no ponderadas cuando no haya diferencias en la intensidad de las relaciones de la red, o cuando esta intensidad no sea un factor relevante.

Por ejemplo, en una red de conocimientos entre familiares, será complicado determinar quién conoce más a quién. En el caso de los medios sociales, nos encontramos con casos como las amistades en Facebook y LinkedIn, o los seguimientos en Twitter: puedes ser

amigo o seguir a alguien una sola vez, no puedes establecer varias relaciones de amistad con una sola persona, ni seguirla más de una vez de manera simultánea.

En una red de relaciones ponderadas, las relaciones toman el valor de 1 cuando existen, y el de 0 cuando no existen.

- **Relaciones ponderadas.** Un conjunto de relaciones podrán considerarse ponderadas cuando consideremos que las diferencias en la intensidad de las relaciones son un aspecto importante que se debe tener en cuenta.

Por ejemplo, no es lo mismo deber a alguien 20 € o 2.000 €, al igual que no es lo mismo mencionar a una persona o llamarla tres veces al año o treinta al día. En Twitter, obviamente, el número de veces que se menciona a alguien es importante. También es importante, por ejemplo, tener en cuenta el número de veces que un usuario interactúa con una página de Facebook.

Según el tipo de aristas, según su direccionalidad y su peso, es posible que nos situemos en cuatro escenarios diferentes de análisis de los datos (ved tabla 1).

Determinar la **dirección** y la **ponderación** de una red es muy importante, puesto que en función de esto se calcularán las métricas de una forma o de otra.

Tabla 1. Tipos de red en función del tipo de aristas

↓ Dirección	Ponderación →	Sin ponderación Todas las aristas pesan igual	Con ponderación Las aristas tienen diferentes pesos
Sin dirección Las aristas no tienen dirección		Red no dirigida y no ponderada (por ejemplo, red de amistades en Facebook o LinkedIn)	Red no dirigida y ponderada (por ejemplo, red de páginas de Facebook según seguidores compartidos)
Con dirección Las aristas sí que tienen dirección		Red dirigida y no ponderada (por ejemplo, red de seguimientos en Twitter o de <i>swipes</i> en Tinder)	Red dirigida y ponderada (por ejemplo, red de menciones en Twitter o de «me gusta» en Instagram)

Fuente: elaboración propia

2.3. Redes de uno y dos modos

Además de las aristas, también es importante fijarnos en las **características** de los nodos de una red. Básicamente hay dos escenarios: cuando todos los nodos son del mismo tipo o cuando hay más de un tipo de nodo en la red.

1) Si todos los nodos de una red pertenecen a la **misma categoría** (por ejemplo, todos son personas, o todos son cuentas de Twitter, o todos son empresas, etc.), estaremos ante una **red de un modo**.

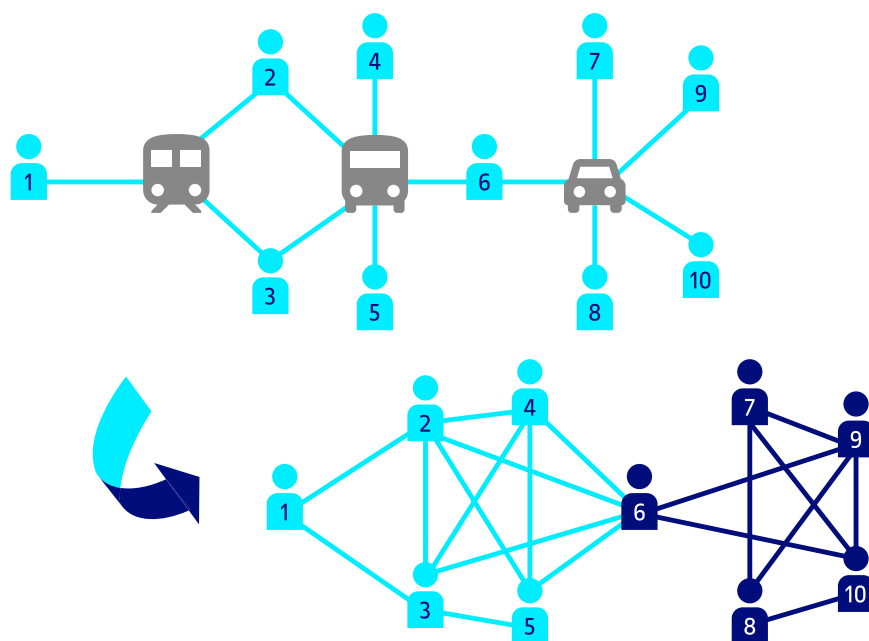
2) En cambio, si los nodos pertenecen a dos **categorías diferentes** (por ejemplo, personas y productos, cuentas y *hashtags*, empresas y estados) estaremos ante una **red de dos modos o bipartita**.

La mayoría de los algoritmos de ARS se aplican en redes de un modo y, por lo tanto, una operación habitual será **trasponer redes de dos modos a redes de un modo**: por ejemplo, si queremos agrupar un grupo de usuarios en función de su medio de transporte preferido (ved figura 3).

Transposición de una red bipartita a una red de un modo

En la figura 3 podemos ver cómo es posible transitar de una red bipartita a otra de un modo. Según los medios de transporte que prefieren una serie de individuos: metro, autobús y coche, hemos podido sintetizar una red trazando una arista entre aquellas personas que comparten, como mínimo, un medio de transporte. El resultado que hemos obtenido es una red que permite dividir la población en dos clústeres: la gente que prefiere los medios públicos (color azul claro) y la gente que prefiere el coche privado (color azul oscuro). Más adelante, veremos cómo funcionan los algoritmos de partición de la red.

Figura 3. Transposición de una red bipartita a una red de un modo



Fuente: elaboración propia

Para llevar a cabo una transposición de una red de dos modos a una red de un modo, debemos hacer una **multiplicación matricial**. Es un procedimiento que implica un cierto grado de complejidad matemática y, sobre todo en conjuntos de datos masivos, puede ser difícil de implementar en un software de cálculo, como por ejemplo MS Excel. El software específico de ARS está preparado para llevar a cabo transposiciones de redes de dos modos a redes de un modo, sea con funciones nativas (por ejemplo, Pajek, UCINET) o con conectores (*plug-ins*) desarrollados por la comunidad (por ejemplo, Gephi). Eso sí, se trata de operaciones que pueden tardar bastante tiempo en computarse completamente en redes masivas.

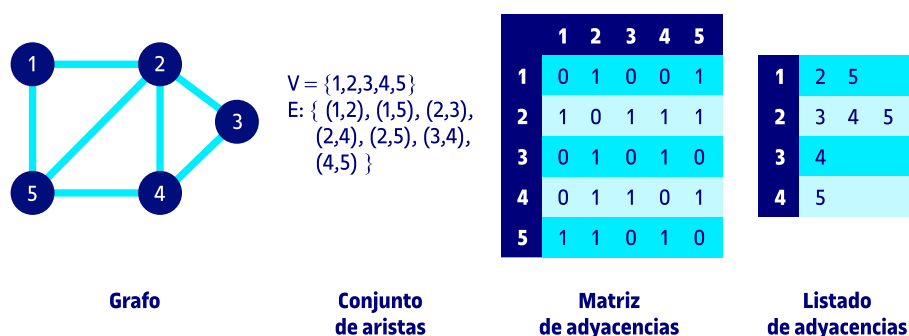
2.4. El lenguaje de las redes

Ya hemos visto que las redes se pueden representar mediante grafos de nodos. Esta es, sin duda, su visualización más eficiente para un analista. En cambio, desde el punto de vista de un ordenador, hay otras representaciones mejores.

Algebraicamente, que es la manera como un ordenador leerá un grafo, los grafos se pueden representar de diferentes maneras. Las más habituales son los conjuntos de aristas, la matriz de adyacencias y el listado de adyacencias (ved figura 4).

- El **conjunto de aristas** es la forma algebraica más sencilla para representar una red. Consiste en un listado de vértices (esto es, V) y de pares de nodos conectados (esto es, E). Con pequeñas variaciones, se trata del formato más habitual que utilizan los programas de ARS y sus principales tipos de ficheros: net, gdf, dl, vna, gml, gexf y graphml.
- La **matriz de adyacencias** es una tabla de doble entrada en la que cada columna y cada fila representan un nodo, y cada intersección, una arista. Las intersecciones toman valores en función de la intensidad del vínculo. En una red no ponderada y no dirigida, los valores serán de 0 o 1 y serán simétricos. La diagonal siempre tomará el valor 0, salvo que haya autorreferencias en el grafo, como por ejemplo una mención de un usuario de Twitter a sí mismo.
- El **listado de adyacencias** es un listado de las conexiones para cada nodo. En el caso de una red no dirigida, no se duplicarán las aristas como en la matriz de adyacencias. Puede ser un sistema más complicado de anotar que los anteriores, pero presenta ventajas importantes en el mundo de las redes digitales, como cuando hemos de hacer un listado de menciones contenidas en una publicación.

Figura 4. Principales formas de representar una red



Fuente: elaboración propia

Todos los formatos anteriores presentarán pequeñas variaciones en función de si la red es dirigida o no dirigida, y también en función de si es ponderada o no ponderada. Gracias al software de ARS y a su orientación hacia perfiles diversos y no necesariamente técnicos, hoy en día no es imprescindible tener un gran dominio de la matemática de conjuntos o de la ciencia computacional para poder llevar a cabo análisis de redes. Ahora bien, sí que es deseable que el analista o la analista esté familiarizado con los formatos de datos que admite el software que utiliza y los formatos en los que puede exportar los resultados de un análisis.

3. Principales métricas en el análisis de redes sociales

Reunir todas y cada una de las métricas que existen en el ARS sería una tarea titánica. En la década de los noventa del siglo XX, los analistas de redes Stanley Wasserman y Katherine Faust (1994) fueron capaces de identificar más de cuatrocientas propuestas para la medida de distintos elementos formales de las redes. Actualmente, el listado todavía sería más amplio. En este apartado, nos limitaremos a presentar algunos de estos elementos, siempre que se den los dos criterios siguientes:

- 1) que sean métricas **disponibles** en el software más habitual (por ejemplo, Gephi o Pajek), y que
- 2) sean métricas **útiles** y aporten valor a los medios sociales.

En consecuencia, el listado de métricas que veremos a continuación no es exhaustivo, pero sí que es capaz de caracterizar una red de manera razonablemente completa.

Las métricas en el ARS pueden reflejar propiedades de los nodos (es decir, métricas nodales) o propiedades de la red en su conjunto (es decir, métricas de grafo):

- 1) **Métricas nodales:** propiedades básicas de los nodos que derivan de sus conexiones con el resto de los nodos o de algoritmos complejos que utilizan estas propiedades básicas. Con las métricas nodales, somos capaces de establecer clasificaciones de usuarios o de categorizarlos en grupos.
- 2) **Métricas de grafo:** métricas que tienen en cuenta las características de los nodos y las aristas de la red en su conjunto y que permiten caracterizar la red completa. Con las métricas de grafo, seremos capaces de ver cuestiones generales que tienen que ver, por ejemplo, con las formas de cohesión de una red y con sus distancias.

3.1. Métricas de los nodos de una red

Ya hemos visto antes que uno de los elementos más esenciales de una red es la arista, la relación entre dos nodos que refleja o representa una interacción contenida en una base de datos.

Referencia bibliográfica

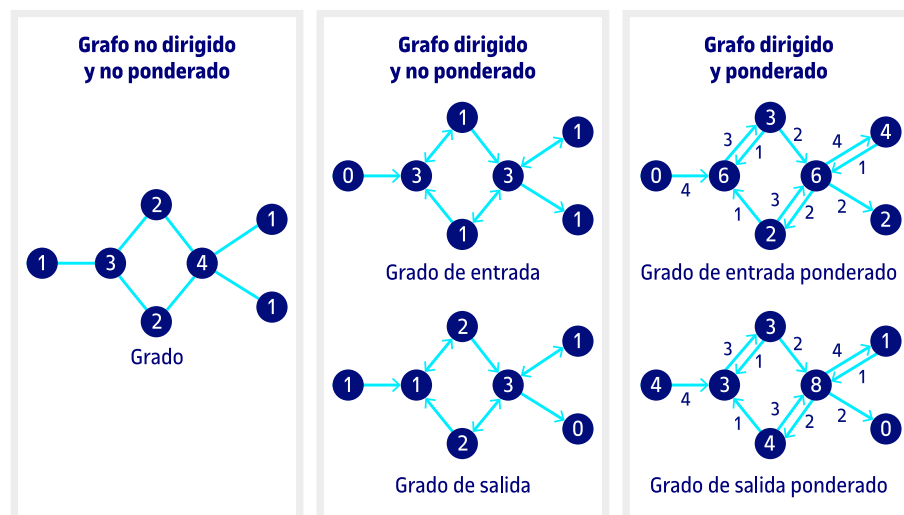
S. Wasserman; K. Faust (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

En función de las aristas de una red y de su distribución, los diferentes nodos tomarán distintos valores para distintos criterios. De estos, el criterio más básico es la **centralidad de grado**, que corresponde al número de vecinos de cada nodo en un grafo.

En una red dirigida, deberemos repartir el grado de un nodo entre su grado de entrada y el de salida –es decir, tendremos que separar las conexiones recibidas de las emitidas–, y en una red ponderada deberemos calcular el grado de manera ponderada, y se considera el peso de las aristas.

Tal y como podemos ver más abajo (ved figura 5), las decisiones que hemos tomado antes sobre la direccionalidad y la ponderación de una red determinan totalmente los cálculos de centralidad de grado.

Figura 5. Métricas básicas de centralidad de grado según el tipo de red



Fuente: elaboración propia

La asunción básica que hay detrás del cálculo de la centralidad de grado es que el número de contactos de un nodo es importante para determinar su poder dentro de la red. Según esta lógica, los nodos más conectados son los más importantes. No obstante, es posible que en nuestro modelo de análisis lo más importante no sea identificar los nodos más conectados, sino aquellos que están mejor conectados.

El ARS es capaz de dar diferentes respuestas a la pregunta sobre qué significa **estar «bien» conectado** en una red. Veamos algunos ejemplos.

1) **Centralidad de intermediación:** número de veces que un nodo está en medio de la ruta más corta (es decir, geodésica) entre todos los pares de nodos de la red (Freeman, 1977). La métrica asume que los nodos capaces de poner en contacto los nodos desconectados son más poderosos.

Un ejemplo de centralidad de intermediación sería cualquier negocio basado en la intermediación de bienes y servicios, como por ejemplo los *marketplaces* de Amazon o Za-

lando, puesto que su modelo de negocio consiste en hacer de intermediarios entre los productores y los compradores.

2) Centralidad de proximidad: distancia geodésica media entre un nodo y los demás nodos de la red (Bavelas, 1950). Esta métrica asume que cuanto menor es el esfuerzo para llegar a cualquier punto de la red, mayor es el poder de un nodo.

Un ejemplo de centralidad de proximidad sería un comercial con una gran cartera de clientes: cuanto menos llamadas tenga que hacer para llegar al director o directora general de una empresa, más valor tendrá este comercial para su propia empresa.

Como sucede con la intermediación, debemos calcular todas las rutas geodésicas de la red para sintetizar la centralidad de proximidad, lo que puede implicar tiempos de computación elevados en datos masivos.

3) Centralidad de vector propio: algoritmo que establece una jerarquía de prestigio entre los nodos de una red en función de si están conectados a otros nodos prestigiosos (Landau, 1895). A los nodos que reciben enlaces provenientes de nodos con un grado de entrada elevado se les imputa una centralidad de vector propio elevada, y la operación se repite en tantas iteraciones como determine el analista (por ejemplo, cien iteraciones). La idea de esta métrica es que un nodo será importante si está vinculado a otros nodos importantes.

Como ejemplo práctico de centralidad de vector propio, podemos pensar en el algoritmo PageRank creado por Google, que es una variante de la centralidad de vector propio que persigue identificar los sitios web con mayor prestigio para indexarlos en las primeras posiciones. De este modo, Google consiguió que sus búsquedas fueran más eficientes que las realizadas con otros buscadores con los que competía a principios de siglo.

4) Hubs y autoridades. Otra variante de la centralidad de vector propio es el algoritmo HITS (Kleinberg, 1999), que ejecuta de manera iterada una serie de procesos que permiten identificar los mejores nodos de la red, atendiendo a dos criterios diferentes:

a) Identifica los mejores *hubs*, que son aquellos nodos que emiten enlaces hacia las mejores autoridades de la red.

b) Identifica las mejores autoridades, que son aquellos nodos que reciben enlaces de los mejores *hubs* de la red.

De este modo, se incorpora un matiz importante que tiene que ver con el rol que desarrolla el nodo en la red: la autoridad prescribe, mientras que el *hub* difunde las prescripciones. Esta métrica es especialmente útil para las redes sociales porque permite identificar a los usuarios que crean contenidos propios y a quienes difunden mediante terceros. Se trata, además, de un algoritmo de muy rápida computación, si se compara con la centralidad de intermediación o la centralidad de proximidad.

Referencia bibliográfica

A. Bavelas (1950). «Communication Patterns in Task-Oriented Groups». *The Journal of the Acoustical Society of America* (vol. 6, núm. 22, págs. 725-730).

Referencia bibliográfica

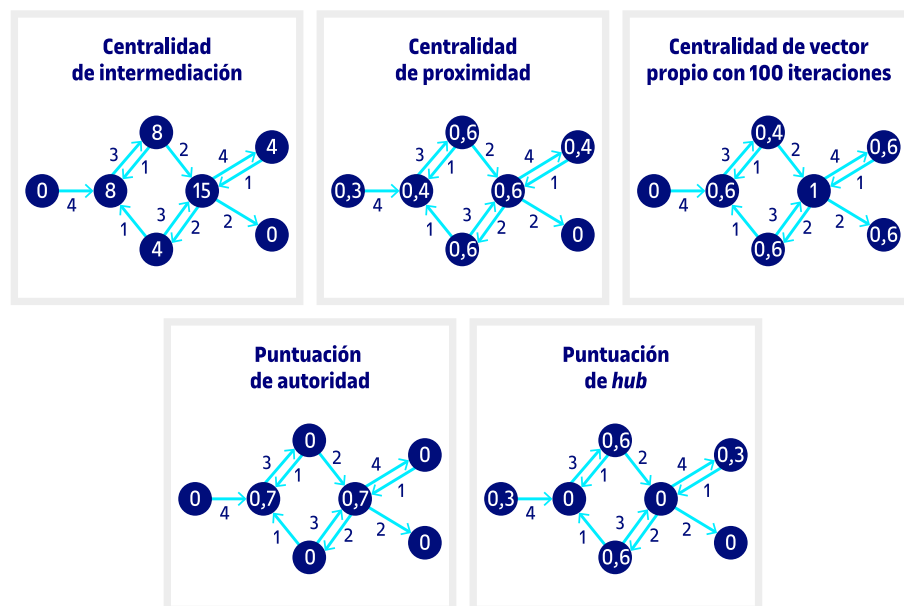
E. Landau (1895). «Zur relativen Wertbemessung der Turnierresultate». *Deutsches Wochensach* (núm. 11, págs. 366-369).

Referencia bibliográfica

J. M. Kleinberg (1999). «Authoritative Sources in a Hyperlinked Environment». *Journal of the ACM (JACM)* (vol. 5, núm. 46, págs. 604-632).

En la comparativa de la figura 6 podemos ver los nodos del mismo grafo, dirigido y ponderado, en el que constan sus puntuaciones según las métricas anteriores. Cada una de las diferentes medidas avanzadas de centralidad aporta información y pueden ser utilizadas de manera complementaria en un análisis. Ahora bien, esto no significa que estemos obligados a utilizar todas estas métricas en un mismo análisis. De hecho, es recomendable elegir las en función de qué es lo más importante que se quiere identificar o explicar y cuál es el tipo de audiencia a la que se dirige el análisis.

Figura 6. Cinco métricas de centralidad para una red dirigida y ponderada



Fuente: elaboración propia

Todas las métricas anteriores aportan información del nodo que permite establecer clasificaciones de varios tipos de centralidad. Se trata, al fin y al cabo, de las variables numéricas asociadas a los nodos y que permiten ordenarlos según determinados criterios. Hay un segundo grupo de variables nodales que no se pueden entender de este modo porque, a pesar de que pueden tomar valores numéricos, son **variables cualitativas**. Se trata de las métricas que dibujan la partición de un grafo y que buscan la caracterización de los subconjuntos de nodos que hay en una red.

1) **Componentes conexos.** En un grafo dirigido, un conjunto de nodos está débilmente conectado (es decir, componentes débilmente conexos) si es posible recorrerlos secuencialmente de manera no repetitiva. Un conjunto de nodos está fuertemente conectado (es decir, componentes fuertemente conexos) cuando, además de la condición anterior, no hay ningún nodo que constituya el final del recorrido, sino que se puede recorrer cíclicamente (Tarjan, 1972). En la teoría de grafos, se denomina **semirruta** a la secuencia de nodos de un componente débilmente conectado, y se denomina **ruta** o **ciclo** a la secuencia de un componente fuertemente conectado. Los componentes fuertes, conec-

Referencia bibliográfica

R. Tarjan (1972). «Depth-First Search and Linear Graph Algorithms». *SIAM Journal on Computing* (vol. 2, núm. 1, pág. 146-160).

tados por rutas, tienden a ser más pequeños y cohesivos que los componentes débiles, conectados por semirrutas. La presencia de muchos componentes fuertes en un grafo proporcionará poder a los buenos mediadores.

2) **Algoritmo Louvain.** Se trata de un algoritmo de aprendizaje no supervisado creado durante los últimos años (Blondel y otros, 2008) que ha resultado la manera más popular para la identificación de comunidades, grupos o clústeres en un grafo masivo. El algoritmo se ejecuta en varias fases y se orienta hacia la **optimización de la modularidad**, una métrica que sirve para medir la calidad de la partición comunitaria de una red (es decir, la calidad de las comunidades que segmentan el grafo) y que veremos más adelante. El algoritmo lleva a cabo agrupaciones aleatorias entre nodos, que evalúa permanentemente el efecto en la modularidad y prescinde de aquellas agrupaciones que van en su detrimento. Como resultado, obtenemos una serie de comunidades diferenciadas y un valor de modularidad que nos ayuda a interpretar la calidad de la partición. Pajek permite la ejecución multinivel del algoritmo Louvain –lo calcula varias veces y devuelve la partición con mejor Q –, mientras que Gephi solo permite la ejecución simple. En ambos casos, hay que determinar manualmente la resolución del algoritmo: una menor resolución propiciará comunidades más pequeñas.

Algoritmos de aprendizaje no supervisado

Los algoritmos de aprendizaje no supervisado son algoritmos capaces de identificar patrones en los datos sin partir de un entrenamiento *ad hoc* implementado por el analista, que nos permiten descubrir propiedades que a simple vista permanecen ocultas.

Las dos métricas anteriores persiguen la caracterización de las partes de un grafo (ved figura 7), pero cada una pone el foco en un factor diferente.

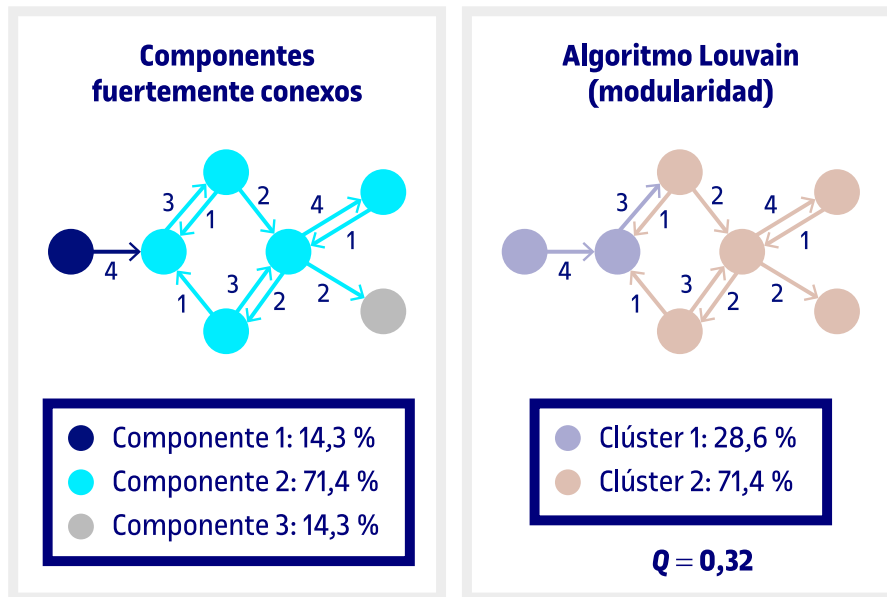
El cálculo de los **componentes débiles y fuertes** busca enfatizar el nivel y los modos de cohesión entre los nodos de una red, mientras que el **algoritmo Louvain** pretende generar grupos en función de criterios menos estrictos, que permite que los nodos inconexos y mutuamente inaccesibles formen parte de un mismo clúster siempre que compartan conexiones significativas con nodos comunes.

El algoritmo Louvain tiende a fallar en la identificación de grupos muy pequeños y cohesivos, mientras que el cálculo de componentes tiende a fallar en la detección de grupos medios. La utilización de uno u otro criterio de partición de un grafo dependerá, una vez más, del tipo de preguntas a las que el o la analista quiera dar respuesta.

Referencia bibliográfica

V. D. Blondel; J. L. Guillaume; R. Lambiotte; E. Lefebvre (2008). «Fast Unfolding of Communities in Large Networks». *Journal of Statistical Mechanics: Theory and Experiment* (vol. 10, P10008).

Figura 7. Dos métricas de partición para una red dirigida y ponderada



Fuente: elaboración propia

3.2. Métricas de la red en su conjunto

Hasta ahora, nos hemos centrado en las métricas de los nodos de la red. Hemos visto métricas cuantitativas y ordinales (por ejemplo, grado de entrada, centralidad de intermediación, centralidad de vector propio), y también cualitativas y categóricas (por ejemplo, componentes fuertemente conectados, particiones del algoritmo Louvain). En este apartado, nos fijaremos en una serie de métricas que nos pueden ayudar a caracterizar la red en su conjunto, en términos de cohesión y distancia.

Una de las cuestiones más elementales en una red es determinar cuál es su **nivel de cohesión**. Se trata, en otras palabras, de identificar hasta qué punto la red se puede considerar un «todo» unificado o cuál es el nivel de saturación relacional que se da.

Hay varias maneras de medir o de evaluar la cohesión de un grafo o de sus partes o clústeres. Cada una de estas métricas nos aporta información complementaria.

1) **Densidad.** La métrica global más sencilla de implementar e interpretar en un grafo es su densidad. La densidad corresponde al número de aristas expresado como la proporción (por ejemplo, de 0 a 1) del máximo posible, considerando que todos los nodos de un grafo podrían estar vinculados. Para calcular la densidad de una red, lo más habitual es considerarla como no ponderada. Entre las limitaciones más importantes de esta métrica se halla su incapacidad

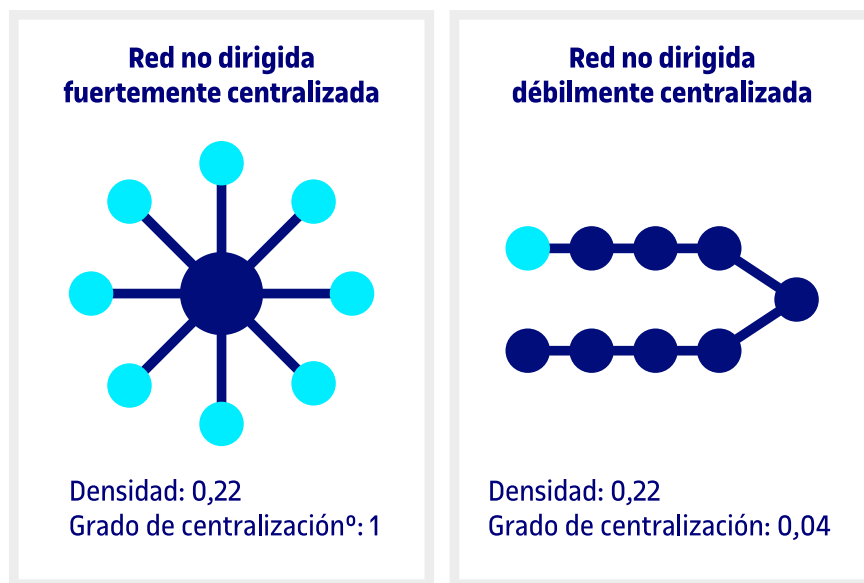
para evaluar la igualdad o la desigualdad en la distribución de las aristas en un grafo, lo que afecta enormemente a la interpretación del modo de cohesión desde un punto de vista cualitativo.

2) Centralización. Se trata de un conjunto de cálculos que se pueden aplicar a todas las métricas de centralidad (por ejemplo, grado, intermediación, proximidad, etc.) que permiten identificar si el poder de los nodos se distribuye homogéneamente en una red o no, mediante la comparación de la distribución observada con la distribución que habría en una red totalmente horizontal con el mismo número de nodos y aristas (Freeman, 1978). La fórmula proporciona un valor que oscila entre 0 (es decir, la descentralización total) y 1 (es decir, la centralización absoluta).

La densidad y la centralización nos hablan de la cohesión en la red. Cada una de estas enfatiza un factor diferente. Veámoslo con el ejemplo de las redes no dirigidas y no ponderadas de la figura 8.

Densidad y centralización

Figura 8. Densidad y centralización en una red no dirigida y no ponderada



Fuente: elaboración propia

Ambas redes tienen 9 nodos y 8 aristas (es decir, 8 aristas de las 36 posibles entre 9 nodos). En consecuencia, las dos tienen la misma densidad, que es de 0,2. Ahora bien, desde el punto de vista de la centralización de grado, las dos redes son muy diferentes: en la primera (esto es, un 100 % de centralización) un solo nodo concentra las conexiones con los demás y, en la segunda (es decir, un 4 % de centralización), los enlaces se distribuyen de manera prácticamente homogénea. En los medios sociales, y si se tratara de redes dirigidas, el primer caso podría corresponder a una red de retuits (es decir, muchos usuarios retuitean uno solo) y, el segundo, a una cadena viral en la que cada usuario cita al anterior. Las métricas de centralización se pueden calcular con Pajek y solo para las relaciones no ponderadas.

Hay un segundo grupo de métricas de cohesión que sirven para evaluar en qué medida los nodos tienden a agruparse en función de algunos de sus **atributos nodales**. Este es uno de los casos en los que combinamos la información relacional propia del ARS con los atributos de los nodos que establecemos externamente (por ejemplo, su género, grupo de edad, nacionalidad, etc.). Aquí no

se trata tanto de identificar el nivel o los modos de cohesión de la red, sino de poder averiguar cuáles son los atributos de los nodos que orientan la cohesión de la red. Este tipo de análisis es muy interesante y útil, si se dispone de datos suficientes sobre todos los nodos (por ejemplo, cuando conocemos su género, oficio, cohorte, aficiones, etc.). Cuando trabajamos con datos de los medios sociales, muchas veces no partiremos de datos tan completos y, por ello, una estrategia habitual será compaginar el ARS con algoritmos de aprendizaje supervisado y no supervisado que permitan generar variables según nombres, descripciones de los usuarios y publicaciones.

1) **Homofilia.** El índice E-I busca identificar si las relaciones en una red son homofílicas o heterofílicas (Krackhardt y Stern, 1988). La operación consiste en evaluar si los nodos tienden a relacionarse, según sus atributos, con aquellos que son iguales (es decir, **homofilia**) o con aquellos que son diferentes (es decir, **heterofilia**). El resultado es un número que oscila entre el -1 (es decir, la homofilia perfecta: todos los nodos solo se relacionan con nodos iguales a estos) y el 1 (es decir, la heterofilia perfecta: todos los nodos solo se relacionan con nodos diferentes a estos). Los valores próximos a 0 indican que el atributo nodal en cuestión no es importante en el establecimiento de las relaciones. La principal limitación del índice E-I es que solo admite cálculos con dos grupos (esto es, **endogrupo** y **exogrupo**).

2) **Modularidad.** Se trata de una métrica diseñada para identificar hasta qué punto una determinada partición de grafo (es decir, su segmentación en comunidades o clústeres) resulta matemáticamente significativa (Newman, 2006). La modularidad se representa con el valor Q y se calcula comparando la partición de un grafo, en el que cada nodo está asignado a una comunidad, con una red aleatoria con el mismo número de nodos y aristas. Los valores iguales o superiores a 0,3 indican la partición de una red, y las comunidades que hay reflejan asociaciones relevantes y significativas, y serán mejores cuanto más se acerquen a 1. Ya hemos visto antes que la modularidad es la base del algoritmo Louvain de identificación comunitaria, pero, en realidad, se trata de un indicador que nos puede ayudar a determinar si cualquier partición es significativa, es decir, si se da por alguna causa diferente al azar. En contraposición con el índice E-I, hay que destacar que la modularidad es capaz de **evaluar la significación de las particiones** con más de dos clústeres. Como contrapunto, se trata de una métrica ciega ante las relaciones heterofílicas.

La modularidad y el índice E-I, disponibles con Pajek, son dos métricas especialmente adecuadas para la contrastación de hipótesis sobre cuáles son los factores clave a la hora de explicar la **conducta relacional** en una red.

Modularidad e índice E-E

Imaginemos, por ejemplo, que queremos estudiar las relaciones entre una serie de individuos y que tenemos dos hipótesis diferentes sobre qué es lo que orienta su conducta relacional:

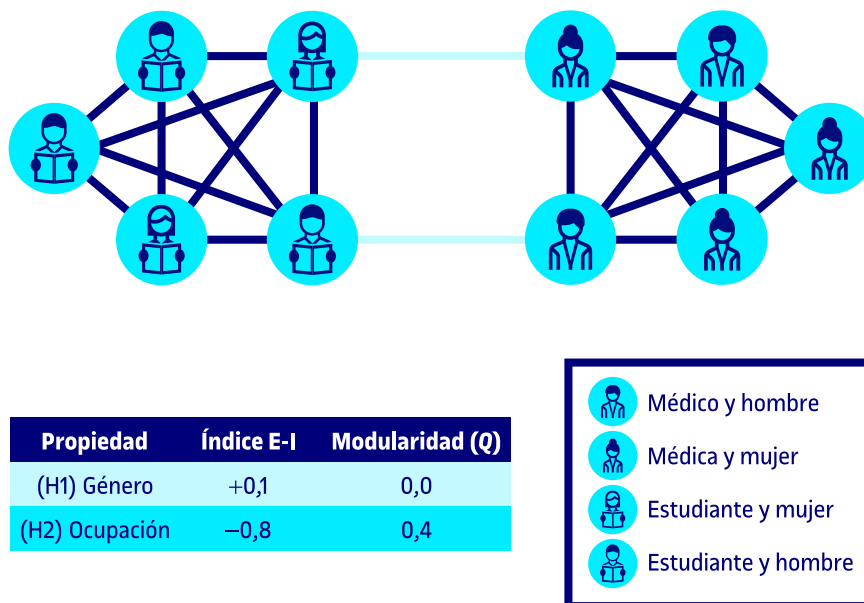
Referencia bibliográfica

D. Krackhardt; R. N. Stern (1988). «Informal Networks and Organizational Crises: an Experimental Simulation». *Social Psychology Quarterly* (págs. 123-140).

- Hipótesis 1 (H1): es el género (es decir, hombres o mujeres) lo que orienta la conducta relacional en la red.
- Hipótesis 2 (H2): es la ocupación (es decir, estudiantes o médicos) la que orienta la conducta relacional en la red.

Tal y como podemos observar en la figura 9, los resultados del análisis en nuestra red serían bastante contundentes: el grupo de individuos en cuestión se relacionaría en función de su ocupación y no de su género. El índice E-I aplicado al atributo nodal de género está próximo a 0 y, por lo tanto, nos indica que los individuos se relacionan entre sí con independencia de su género. En cambio, el índice E-I es de $-0,8$ en cuanto a la ocupación, lo que nos indica que en esta red se da una homofilia muy intensa entre estudiantes, por un lado, y médicos, por el otro. El valor Q nos sugiere que la distribución del género en la red es la misma que encontraríamos en una distribución aleatoria y que, por lo tanto, no es significativa. Sí que resulta importante la distribución según la ocupación, puesto que es significativamente diferente a la que se consigue por azar.

Figura 9. Género y ocupación en un grafo no dirigido y no ponderado



Fuente: elaboración propia

Más allá de la cohesión de la red, el segundo grupo de métricas globales son las métricas de distancias.

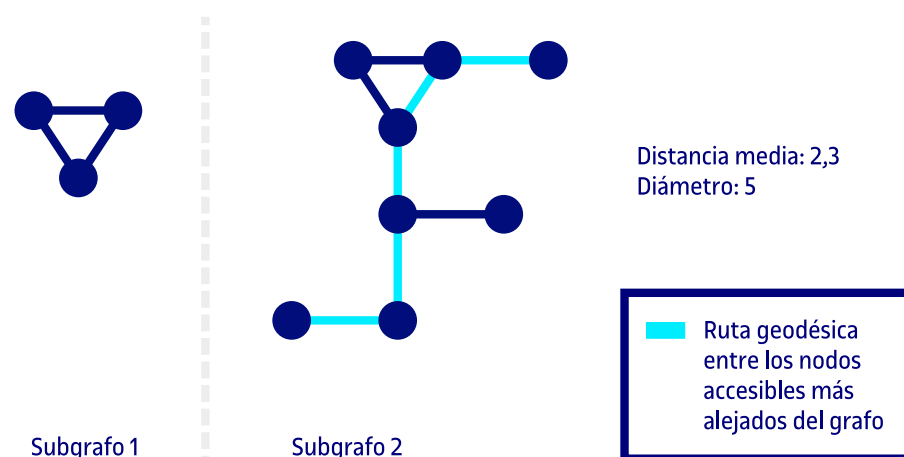
Las **métricas de distancias** son indicadores que nos informan sobre la calidad de los vínculos entre los nodos desde la perspectiva de cómo de eficiente puede ser la difusión en una red.

En este caso, no se trata de evaluar si los nodos están muy o poco interconectados, sino de evaluar si las conexiones establecidas permiten una fácil propagación de lo que constituye el contenido compartido en la red: ya sea información, dinero o un virus.

1) **Distancia media:** número de «saltos» entre los nodos que hay de media en todas las rutas geodésicas entre los pares de nodos de la red. Es decir, la métrica consiste en la media del recuento de aristas que hay entre todos los pares de nodos mediante la ruta más corta entre los dos. Esta es, sin duda, la métrica más famosa del ARS, popularizada por medio de la **teoría de los seis grados de separación**, testada experimentalmente por Stanley Milgram (1967). Estudios más recientes llevados a cabo con la población de Facebook sugieren, en cambio, que la distancia media entre cualquier par de personas del planeta Tierra podría acercarse más a 4 (Backstrom y otros, 2011).

2) **Diámetro:** número de «saltos» entre los nodos que conectan los dos nodos más alejados de la red por la ruta más corta. La métrica, que solo se puede calcular cuando la red no consiste en subgrafos totalmente aislados (figura 10), utiliza la metáfora circular del diámetro, puesto que es la distancia entre los dos puntos más alejados en una circunferencia.

Figura 10. Subgrafos, distancia y diámetro en una red no dirigida y no ponderada



Fuente: elaboración propia

Las métricas globales de distancia que vemos en la figura 10 están relacionadas con las métricas nodales, como por ejemplo la centralidad de intermediación o de proximidad, y tienen en común con estas una gran necesidad computacional. Cuando nos encontramos con una red altamente fragmentada, compuesta de muchos subgrafos, este tipo de métricas perderán valor explicativo, puesto que solo se podrán aplicar a cada uno de estos subgrafos y no a la red en su conjunto. Otra limitación es que puede ser complicado establecer comparaciones entre redes con este tipo de medidas: las distancias dependen del volumen de nodos y de aristas en un grafo y no hay un único criterio para determinar si estamos ante una distancia «corta» o «larga».

Referencia bibliográfica

L. Backstrom; P. Boldi; M. Rosa; J. Ugander; S. Vigna (2011). «Four Degrees of Separation». *Proceedings of the 4th Annual ACM Web Science Conference* (págs. 33-42).

4. Visualización de redes sociales

Hasta ahora, nos hemos centrado en el aspecto matemático y algebraico de los grafos, en sus métricas y en cómo interpretarlas. En este apartado, exploraremos, por un lado, las herramientas de visualización de los grafos y, por otro, veremos las diferentes estrategias para representarlos siguiendo diferentes criterios y haciendo énfasis en algunas de sus propiedades.

4.1. Software de análisis de redes sociales

La historia del ARS ha estado estrechamente vinculada a la tecnología, la inteligencia artificial y la computación. Dado que las redes que se deben analizar congregan unas decenas de nodos y aristas, se hace totalmente imprescindible que el o la analista cuente con un buen soporte en cuanto a software.

Sobra decir que en entornos de datos masivos provenientes de los medios sociales, en los que las bases de datos pueden tener millones de registros, es crucial contar con un **software adaptado y preparado** para ejecutarlas con cálculos y visualizaciones masivas.

El primer software de ARS fue escrito por James S. Coleman y Duncan MacRae a finales de la década de los cincuenta y estaba diseñado para identificar nodos interconectados en una red de hasta mil nodos, lo que implicaba una cantidad de datos absolutamente gigantesca para la época. Durante dos décadas, los analistas de redes desarrollaron varios programas orientados a la resolución de problemas específicos: Socpac, Blocker, Sock, Complt, Catij, Concor, Structure, Negopy, Sonet, Center, Gradap, Cobloc, etc. No fue hasta los años ochenta cuando se empezó a desarrollar un software orientado a la resolución de problemas múltiples (Freeman, 2012). De aquella época, destacan **Sonis** y **Ucinet**.¹

⁽¹⁾Ucinet todavía se utiliza para el análisis de redes no masivas.

Pero el primer software de ARS específico para datos masivos no vino de la tradición científicosocial y computacional norteamericana, sino del otro lado del antiguo telón de acero. El matemático Vladimir Batagelj, de la Universidad de Liubliana, en Eslovenia, presentó en 1996 la primera versión de Pajek. El programa tuvo mucho éxito, inicialmente, entre bioquímicos estudiosos del genoma y de las interacciones proteicas. Hoy en día, **Pajek** sigue siendo uno de los programas más potentes de ARS. Probablemente el más potente de todos en cuanto a su **capacidad de cálculo matemático**. Pajek es gratuito para usos no comerciales y está disponible para Windows, a pesar de que funciona perfectamente en entornos MAC o Linux mediante Wine.

En cambio, otros softwares más nuevos sí que mejoran la **capacidad de visualización de los grafos**. El mejor ejemplo de ello es **Gephi**, un software libre con licencia GNU. Gephi fue desarrollado inicialmente por estudiantes de la Universidad Tecnológica de Compiègne y, actualmente, cuenta con una gran comunidad de desarrolladores y desarrolladoras en GitHub. Gephi funciona en cualquier sistema operativo que admita Java y OpenGL.

Además de Pajek y Gephi, actualmente hay un gran número de programas y librerías de Python y R disponibles para llevar a cabo análisis de redes. Entre otros, hay que destacar Igraph, NetworkX y el conector (*plug-in*) de MS Excel NodeXL, muy popular por su sencillez de uso y por su bajo coste. Las librerías Igraph y NetworkX, en cambio, son totalmente gratuitas y de código abierto.

4.2. Algoritmos de visualización de redes sociales

Un grafo es simultáneamente una herramienta de cálculo matemático y una representación visual de un conjunto de datos relacionales. Por eso, una de las principales características del análisis de redes sociales es su **orientación visual**.

Fundamentalmente, hay dos tipos de diseño o de representación para un grafo:

1) **Representación conceptual**: consiste en distribuir los nodos y las aristas de tal manera que se ajusten a una forma o configuración preestablecida, y que permitan observar uno o varios atributos de los grafos (ved figura 11). Algunas de las representaciones conceptuales más populares son las siguientes:

a) **Representación circular**: los nodos se disponen formando un círculo, ordenados por algún criterio ordinal (por ejemplo, el grado de entrada, la intermediación) y, opcionalmente, pintados en función del clúster de pertenencia. Las representaciones circulares son óptimas cuando se quiere representar una clasificación de nodos, y también visualizar sus relaciones.

b) **Representación radial**: los nodos se ordenan formando un círculo en función de un criterio ordinal, pero después se concatenan todos los nodos con la misma puntuación ordinal, lo que deja en el centro un ranquin de casos únicos. Opcionalmente, también pueden pintarse los nodos en función del clúster de pertenencia. Este tipo de representación pone más énfasis en el volumen de casos parecidos en una red y no tanto en las relaciones entre los nodos.

c) **Representación geográfica**: cada nodo tiene una asignación geográfica y está dispuesto en un mapa. Los nodos pueden ponderarse según un criterio ordinal y pintarse en función de un criterio categórico, como por ejemplo el clúster de pertenencia.

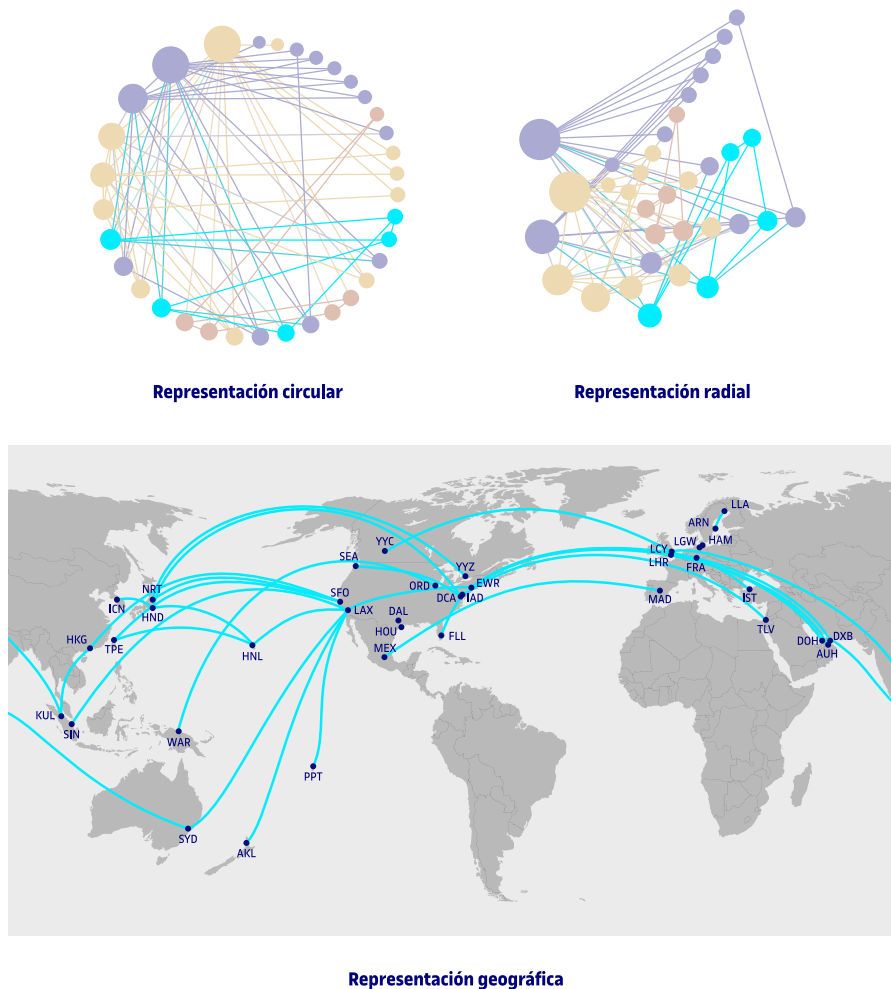
Representación circular, radial y geográfica

La representación circular puede servir, por ejemplo, para representar ránquines y relaciones entre *hashtags*, en función de las veces que han sido más utilizadas conjuntamente.

La representación radial, por ejemplo, puede servir para visualizar el ranquin de mediadores en una red de menciones entre usuarios y para enfatizar cuántos de estos tienen la misma capacidad de intermediación.

Las representaciones geográficas son deseables cuando se quiere expresar la relación entre casos geográficamente ubicados, como, por ejemplo, relaciones de seguimiento entre usuarios de diferentes ciudades o países.

Figura 11. Representaciones conceptuales de redes sociales



Fuente: elaboración propia y adaptación de Wikimedia Commons

2) **Algoritmos complejos:** reflejan varias propiedades de los nodos, las aristas y los grafos (ved figura 12). La gran mayoría de estos algoritmos son **de fuerza bruta**; esto significa que lo que hacen es calcular un gran número de soluciones posibles y solo quedarse con la mejor: la que se ajusta de manera más adecuada a una serie de parámetros preestablecidos. Estos parámetros son diferentes para cada uno de estos algoritmos y dependen de su entrenamiento. Algunos de los algoritmos complejos son los siguientes:

a) **Force Atlas 2:** el algoritmo de fuerza bruta más popular de Gephi, diseñado por Jacomy y sus colaboradores (2014). El algoritmo acerca los nodos que están vinculados con más intensidad y los aleja de aquellos con los que el vínculo es débil o no existe. Se trata de un algoritmo útil en redes dirigidas y no dirigidas, ponderadas y no ponderadas, pero es especialmente interesante para representar redes dirigidas y ponderadas, como por ejemplo las redes de menciones, y resulta muy útil para representar la **semejanza entre los nodos**, en función de su comunidad de pertenencia. Es un algoritmo que funciona muy bien si las comunidades que se representan han sido identificadas con el algoritmo Louvain (Blondel y otros, 2008), puesto que ambos utilizan formas de optimización parecidas. El algoritmo Force Atlas 2 es capaz de representar exitosamente grafos con varios millones de nodos, eso sí, puede requerir bas-

tante tiempo y recursos computacionales. Los algoritmos **Yifan Hu Multilevel** y **Fruchterman-Reingold**, más antiguos y también disponibles con Gephi, tienen un funcionamiento muy similar. No obstante, se trata de algoritmos menos preparados para las redes masivas.

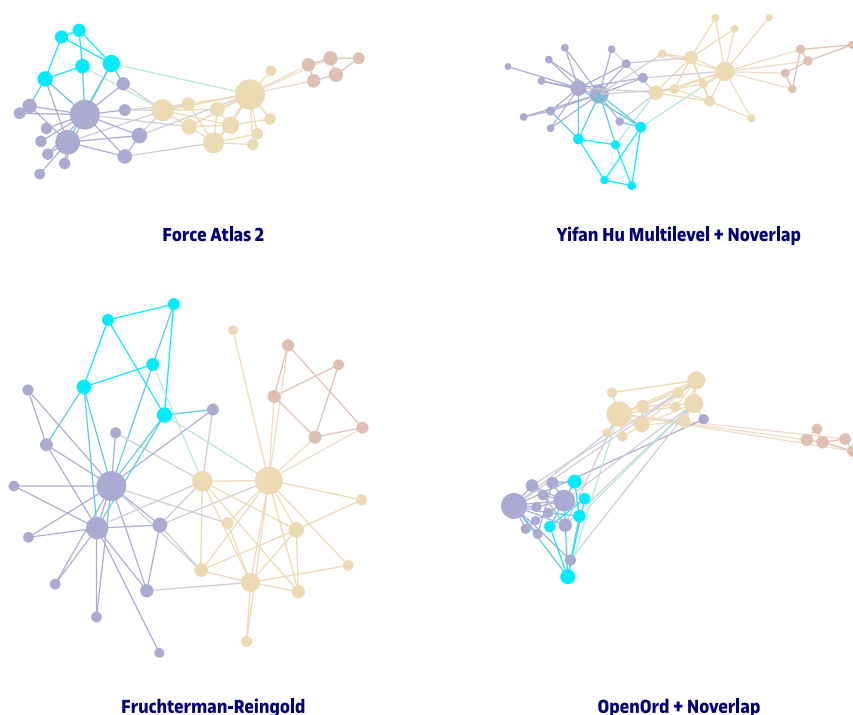
b) **OpenOrd**: un algoritmo diseñado para representar grafos de millones de nodos, igual que Force Atlas 2. Su objetivo principal es **separar los grupos de nodos inconexos**; es decir, persigue algo muy parecido a los algoritmos anteriores, pero enfatiza la diferencia en lugar de la similitud entre los nodos. Su principal ventaja es que permite sintetizar los grafos masivos en tiempos muy breves y con menos recursos computacionales que los algoritmos anteriores. Para conseguirlo, el algoritmo prescinde de las aristas más débiles de la red. En contraposición con Force Atlas 2, es un algoritmo menos preciso y versátil, puesto que solo tiene en cuenta una parte de las relaciones del grafo y está optimizado para redes no dirigidas y ponderadas. Además, su complementariedad con el algoritmo Louvain es menor que la de Force Atlas 2.

c) **Noverlap**: un tipo de algoritmo que tiene la función de asistir a los que no tienen incorporado nativamente el parámetro de **evitar la superposición entre nodos** (por ejemplo, Yifan Hu Multilevel u OpenOrd). El algoritmo Noverlap permite ajustar el radio de separación entre nodos y la velocidad de ejecución. A diferencia de los algoritmos anteriores, para usarlo deberemos haber aplicado antes otro algoritmo que determine la configuración de partida. Si Noverlap se ejecuta a una velocidad elevada, se obtendrá un resultado más alejado de esta configuración de partida. Por el contrario, si se ejecuta más lentamente, la configuración de partida será más respetada y el resultado global del proceso será mejor. Esto puede implicar tiempo de computación muy elevado, sobre todo en redes masivas.

La visualización de los datos constituye un vehículo de transmisión de información muy importante: de los datos al analista y del analista al lector final, que puede ser alguien diferente al analista o no. Sin embargo, es importante entender –y hacer entender al eventual lector final– que una representación visual, como por ejemplo un grafo, es el resultado de un diseño específico y no es recomendable interpretarla solo con la intuición de cada uno.

Es importante que el analista interprete los datos y facilite la interpretación al lector final.

Figura 12. Representaciones complejas de redes sociales

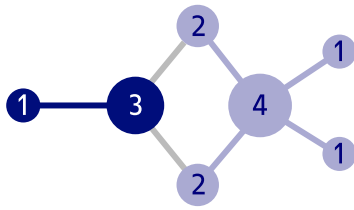


Fuente: elaboración propia

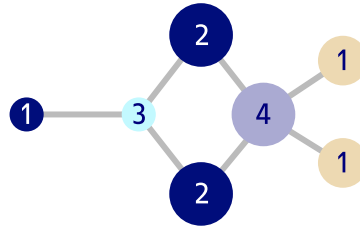
El lector final, a veces, será alguien con un perfil similar al del analista, pero la mayoría de las veces será un lector con competencias y habilidades diferentes: un compañero o compañera de trabajo, el alumnado de un curso o, incluso, un público general sin conocimientos de análisis de datos. Es fundamental que el analista considere este público al que se dirige en todo momento a la hora de generar visualizaciones eficientes y capacitadas para transmitir exactamente lo que pretende, mediante la interpretación que hará de estas.

Cualquier visualización debe servir para responder, como mínimo, a una pregunta clave. Las preguntas clave a las que trataremos de responder con un grafo podrán tener que ver con aspectos como, por ejemplo, el liderazgo en la red, su composición comunitaria o su distancia. Así, podremos recurrir a visualizaciones que enfatizan el grado de entrada cuando queramos identificar a los principales líderes de una red de retuits, o a una visualización que enfatice la distribución comunitaria de una conversación, si queremos identificar los diferentes grupos de usuarios que la componen. Para responder correctamente a estas preguntas, es importante utilizar los parámetros de diseño de manera lógica y congruente, por ejemplo, ponderando la medida de los nodos en función de las variables ordinales y no categóricas, y pintarlos en función de las variables categóricas y no ordinales (ved figura 13).

Figura 13. Buenas y malas prácticas en la representación de variables en un grafo



Medida: grado
Color: clúster



Medida: clúster
Color: grado



Fuente: elaboración propia

Bibliografía

Backstrom, L.; Boldi, P.; Rosa, M.; Ugander, J.; Vigna, S. (2012). «Four Degrees of Separation». *Proceedings of the 4th Annual ACM Web Science Conference* (págs. 33-42).

Bavelas, A. (1950). «Communication Patterns in Task-Oriented Groups». *The Journal of the Acoustical Society of America* (vol. 6, núm. 22, págs. 725-730).

Blondel, V. D.; Guillaume, J. L.; Lambiotte, R.; Lefebvre, E. (2008). «Fast Unfolding of Communities in Large Networks». *Journal of Statistical Mechanics: Theory and Experiment* (vol. 10, P10008).

Deutsch, K. W. (1952). «On Communication Models in the Social Sciences». *Public Opinion Quarterly* (vol. 3, núm. 16, págs. 356-380).

Granovetter, M. S. (1973). «The Strength of Weak Ties». *American Journal of Sociology* (vol. 6, núm. 78, págs. 1360-1380).

Euler, L. (1736). «Mechanica Sive Motus Scientia Analytica Exposita. Instar Supplementi ad Commentaria». *Acad. Scient. Imper* (vol. 1). Ex Typographia Academiae Scientiarum.

Freeman, L. C. (1977). «A set of Measures of Centrality Based on Betweenness». *Sociometry* (págs. 35-41).

Freeman, L. C. (1978). «Centrality in Social Networks Conceptual Clarification». *Social Networks* (vol. 3, núm. 1, págs. 215-239).

Freeman, L. C. (2012). *El desarrollo del análisis de redes sociales: un estudio de sociología de la ciencia*. Bloomington, Indiana: Palibrio.

Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. (2014). «ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software». *PloS One* (vol. 6, núm. 9, e98679).

Kleinberg, J. M. (1999). «Authoritative Sources in a Hyperlinked Environment». *Journal of the ACM (JACM)* (vol. 5, núm. 46, págs. 604-632).

Krackhardt, D.; Stern, R. N. (1988). «Informal Networks and Organizational Crises: an Experimental Simulation». *Social Psychology Quarterly* (págs. 123-140).

Landau, E. (1895). «Zur relativen Wertbemessung der Turnierresultate». *Deutsches Wochenschrift* (núm. 11, págs. 366-369).

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

Milgram, S. (1967). «The Small World Problem». *Psychology Today* (vol. 1, núm. 2, págs. 60-67).

Newman, M. E. (2006). «Modularity and Community Structure in Networks». *Proceedings of the National Academy of Sciences* (vol. 23, núm. 103, págs. 8577-8582).

Tarjan, R. (1972). «Depth-First Search and Linear Graph Algorithms». *SIAM Journal on Computing* (vol. 2, núm. 1, págs. 146-160).

Wasserman, S.; Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

White, H. C.; Boorman, S. A.; Breiger, R. L. (1976). «Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions». *American Journal of Sociology* (vol. 4, núm. 81, págs. 730-780).

