



CAPTURA DE ESTADOS MENTALES EN REDES SOCIALES

Una filosofía de la mente extendida

Capture of Mental States in Social Networks. A Philosophy of the Extended Mind

KARIM GHERAB MARTÍN

Universidad Rey Juan Carlos, España

KEY WORDS

*Philosophy of Mind
Intentionality
Internet
Psyche
Social Networks
Digital Technology*

ABSTRACT

This article analyzes how technology multinationals, and in particular those that are dedicated to social networks, are developing techniques to build mental metrics of users who browse the Internet in order to know their emotions, their feelings, their tastes, their phobias, and, ultimately, their psyche. Some simple examples of how these techniques can be implemented are shown. Specifically, we will focus on a very particular trait of the mental: intentionality.

PALABRAS CLAVE

*Filosofía de la mente
Intencionalidad
Internet
Psique
Redes sociales
Tecnología digital*

RESUMEN

Este artículo analiza cómo las multinacionales tecnológicas, y en particular aquellas que se dedican a las redes sociales, están desarrollando técnicas para construir métricas mentales de los usuarios que navegan por internet con el fin de conocer sus emociones, sus sentimientos, sus gustos, sus fobias, y, en definitiva, su psique. Se muestran algunos sencillos ejemplos de cómo se pueden implementar estas técnicas. En concreto, nos centraremos en un rasgo muy particular de lo mental: la intencionalidad.

1. Introducción

Es frecuente que matemáticos, físicos, químicos, biólogos, ingenieros e incluso economistas y sociólogos utilicen la Teoría de Grafos¹ para desarrollar teorías que tienen que ver con redes de todo tipo: redes moleculares, redes neuronales, redes de computadores, redes de personas, redes comerciales, redes sociales, redes de organizaciones criminales, etc. El elemento común a todas estas redes tangibles es que todas ellas son representables idealmente como nodos y aristas, donde los nodos pueden ser moléculas, neuronas, computadores, personas, instituciones, o incluso redes de redes. Naturalmente, del mismo modo que los nodos representan entidades, las aristas representan relaciones entre estas entidades, y estas relaciones (o aristas) pueden tener propiedades diversas: pueden ser unidireccionales o bidireccionales (en cuyo caso la arista se convierte en una flecha, con una o dos puntas), puede haber una ponderación de intensidades (lo cual puede representarse con el grosor de la arista), puede haber una ligera distinción cualitativa entre ellas (lo cual puede representarse con aristas de distintos colores), etc.

Lo que no es frecuente es imaginar aristas que representen procesos mentales entre nodos, y aún menos actitudes intencionales. El presente artículo tratará sobre la interpretación de las relaciones/aristas entre nodos como intenciones que se proyectan de unos nodos hacia otros. En particular, veremos cómo las multinacionales tecnológicas, y en particular aquellas que se dedican a las redes sociales, están desarrollando técnicas para construir métricas mentales de los usuarios que navegan por internet con el fin de conocer sus emociones, sus sentimientos, sus gustos, sus fobias... y, en definitiva, su *psique*.

El alcance del presente artículo es limitado y no pretende entrar en detalle en el incipiente mundo de las métricas mentales. Los ejemplos mostrados a continuación son esbozos de la manera en que pueden medirse algunos procesos mentales tales como las intenciones. Hay otros

procesos mentales que pueden medirse gracias a los clics que hacen continuamente los usuarios en la red, pero aquí nos centraremos únicamente en los procesos intencionales de los usuarios.

2. Google como antecedente de las redes sociales

Antes de la irrupción de Google en 1999, los buscadores estaban diseñados como portales. Los buscadores más utilizados eran *Altavista* y *Yahoo!*, que, como puede verse en las imágenes de abajo, eran portales que categorizaban las búsquedas. En otras palabras, tanto *Altavista* como *Yahoo!* mostraban un índice temático con el fin de tutelar a los usuarios en su entrada al nuevo y desconocido mundo de Internet. El índice temático permitía acceder a páginas web previamente seleccionadas por los expertos de *Altavista* y *Yahoo!*, de modo que el usuario sólo tenía que hacer clic para irse adentrando en los subtemas de cada tema hasta encontrar un listado de páginas que los expertos de dichos portales consideraban más relevantes. Este modelo tutelado de acceso a los contenidos de Internet extrapolaba al mundo digital la tradición bibliotecaria de organizar el conocimiento en formato *tesauro*, esto es, en clasificaciones temáticas jerarquizadas y más o menos completas.

Sin embargo, tanto *Altavista* como *Yahoo!* ofrecían otra manera de acceder a los contenidos de la Red: una caja en la que los usuarios podían introducir una palabra o cadena de palabras, y luego hacer clic en el botón 'Search' que estaba a su derecha. Al entrar de este modo, la *caja mágica* devolvía un larguísimo listado desordenado de páginas, donde no parecía haber ningún criterio para separar el grano de la paja, es decir, ordenar las páginas web en función de la calidad de sus contenidos.

Estas cajas mágicas recuperaban y priorizaban las páginas web en función del número de veces que se repitiera la palabra (o cadena de palabras) buscada en dicha página, y otras técnicas muy similares. Es evidente que priorizar las páginas web en función del número de veces que apareciera una palabra no era un buen indicador de la calidad de una página, pero los ingenieros no tenían una manera mejor de destacar lo bueno e ignorar lo malo. Buscaban cantidades, calculaban ratios de frecuencia de significantes, priorizaban

¹ La Teoría de Grafos es la teoría matemática que estudia los modelos reticulares abstractos. Los teóricos de grafos imaginan puntos (nodos) y las posibles *relaciones* entre esos puntos, que suelen representar como una línea recta.

la sintaxis porque no sabían enfrentar la labor de inyectar carga semántica a las búsquedas ni de dar respuesta a la recuperación de conceptos y no de meros términos.

Por eso, era habitual que el usuario poco avezado en realizar búsquedas booleanas con la caja mágica, estuviera abocado irremediabilmente al índice temático preparado por los expertos de *Altavista* y *Yahoo!*, que, ellos sí, sabían hacer búsquedas sintácticas avanzadas mediante el uso de la *lógica de Boole* aplicada a cadenas de palabras.

Y entonces apareció *Google*, con un fondo totalmente blanco (en versión *beta*), indicando con claridad que la búsqueda mediante la utilización de índices temáticos había llegado a su fin. La austera página de *Google* era una clara declaración de intenciones. Venía a decir se habían terminado los portales para hacer búsquedas, porque el nuevo buscador tenía una caja mágica que superaba cualquier índice temático. El *sorpas* de *Google* a *Altavista* y *Yahoo!* no se hizo esperar; los usuarios de internet empezaron a usar *Google*.

Figura 1. Imagen del portal-buscador *Altavista* el 29 de abril de 1999.



Fuente: Recuperado con WayBackMachine en www.archive.com

Figura 2. Imagen del portal-buscador *Yahoo!* el 29 de abril de 1999.



Fuente: Recuperado con WayBackMachine en www.archive.com

Figura 3. Imagen del buscador *Google* el 28 de abril de 1999.



Fuente: Recuperado con WayBackMachine en www.archive.com

La magia de *Google* lograba ordenar las páginas web en función de su importancia. Naturalmente, esa caja no era infalible, pero era infinitamente mejor que nada de lo visto hasta entonces. Y ahí empezó el lento e inexorable declive de *Altavista* y *Yahoo!*. La clave del éxito de *Google* fue el inicio de lo que hoy llamamos Web 2.0, que se basa entre otras cosas en la colaboración P2P, o sea *peer-to-peer* (O'Reilly, 2005): *Google* creó un algoritmo que evaluaba las páginas web en función de los *comportamientos intencionales* de los usuarios en la Red. No había

expertos ordenando jerárquicamente los contenidos, sino que eran los propios usuarios (los pares o *peers*) los que, con sus acciones, estaban trabajando, sin saberlo, para *Google* y, por consiguiente, para sus clientes (otros pares o *peers*). Pero antes de explicar cómo funciona el algoritmo de *Google*, haremos una incursión histórica para conocer su origen. La invención de *PageRank*, así se llama el algoritmo, no surgió de la nada, sino que tiene detrás una historia muy interesante.

3. El origen del algoritmo de *Google*

La idea sobre la que se basa *PageRank* tiene su origen en los EEUU, a finales del siglo XIX. Es frecuente ver en algunas películas estadounidenses alguna escena de un juicio en la que el fiscal o el abogado defensor cita casos anteriores porque sientan precedente. Tal recurso a un caso anterior de sentencia favorable puede servir de poderoso argumento para quien lo trae al caso, de modo que hacen referencia al volumen y a la página en el que dicho caso se encuentra registrado o bien al artículo, capítulo, sección y publicación en el que se encuentra determinado estatuto. Así, nació una publicación en el último cuarto del siglo XIX, *Shepard's Citations*, perteneciente a la empresa consultora Franck Shepard Company, que se encargaba de proveer a los juristas de tales citas (casos, estatutos y sentencias) y, especialmente, de mostrar la *historia* de un caso registrando todos los casos posteriores que lo habían citado, informando así del resultado de la sentencia de los casos posteriores que citaron ese caso. La indexación de los casos jurídicos detallaba incluso, mediante una sencilla codificación por letras el historial de cada caso, indicando con la letra "e" aquellos casos en los que el tribunal dio una explicación del caso original citado, con la letra "a" para denotar que en esa ocasión el tribunal confirmó la validez del precedente, con la letra "d" para especificar que se introdujeron una serie de aclaraciones o distinciones respecto del caso original, etc. Veamos un ejemplo publicado por Adair (1955)²:

	101 Mass	210
	112 Mass	65
e	130 Mass	89
	165 Mass	210
d	192 Mass	69
	205 Mass	113
e	212 Mass	173
	221 Mass	210
a	281 U.S.	63
	35 H.L.R.	76

La explicación de la anterior tabla es la siguiente:

- El caso original a partir del cual el abogado comienza su investigación (para defender a su cliente) es el caso 101 que aparece en la página 210 del apartado de informes correspondientes al estado de Massachussets (EEUU).
- Los casos que aparecen en el listado son todos aquellos casos que citaron el caso original, de modo que:
 - En el caso 130 Mass 89 el tribunal dio una explicación del caso original (denotándolo con el código "e");
 - En el caso 192 Mass 69 el tribunal introdujo una aclaración o distinción ("d") con respecto del caso original limitando el área de validez del caso como precedente;
 - Y en el caso 281 U.S. 63 el Tribunal Supremo de los EEUU aseveró ("a") que el caso representaba un buen precedente como ley;
 - La aseveración Tribunal Supremo de los EEUU fue además publicada en el *Harvard Law Review* (H.L.R.).

La siguiente imagen muestra casos reales, un extracto de una página cualquiera:

² Adair, W. C. (1955). "Citation Indexes for Scientific Literature", *American Documentation* 6, pp. 31-32.

CITATIONS TO LAW, AS PUBLISHED IN COURT DECISIONS														
CITED COURT DECISIONS														
- 171 - 352FS411 So C 1948403	- 714 - 351FS586 6351FS985 57FRD99	- 780 - One 504P241408	- 167 - 471F2d1190 352FS708	- 541 - 471F2d11091 127CaA515 521B2460	- 175 - 31NY951	272Scd1624 Calif	18MdA529 292Min402 96PRR389 96PRR382 96PRR393 96PRR394 96PRR711 108R1531 1TnCr123 1TnCr185 1TnCr350 1TnCr417 17Cr698	- 213 - 352FS741 174C2871 174C2857 71R119 16MdA 72Msc2 17Cr698	- 724 - 352FS661	Vol. 387	- 213 - 470F2d1635 1948E319 NY	- 244 - 340S2d812	- 218 - 470F2d1153 105CaR107 106CaR110 106CaR112 470F2d1380 505P24531 470F2d1718 506P24534 4470P24949 505P24536 471F2d1119 471F2d425	470F2d1380 505P24531 470F2d1718 506P24534 4470P24949 505P24536 471F2d1119 471F2d425

Al leer el artículo de Adair (1955), el químico y lingüista³ convertido a la bibliometría Eugene Garfield vio rápidamente la validez de este esquema para la indexación del conocimiento en la ciencia (Garfield, 1955) y lo aplicó de inmediato a los fondos bibliográficos de patentes en el área de la química (Garfield, 1957). En efecto, el funcionamiento de las oficinas de patentes es similar al esquema mostrado puesto que los examinadores han de asegurarse que el invento es realmente original y recurren a los fondos bibliográficos para confirmarlo. Así, deben respaldar sus conclusiones haciendo referencia a estudios y patentes anteriores de modo que todo el historial de una rama de invención quede registrado con precisión y sea recuperado eficazmente y sin lagunas. Naturalmente, en cualquiera de los tres casos (Derecho, Patentes y Ciencia) es preciso iniciar las investigaciones a partir de un caso, una patente o un artículo concreto, lo cual puede no ser sencillo si no se conoce el área de conocimiento correspondiente. Sin embargo, la dificultad es superable en cuanto alguien nos indique un artículo por el que comenzar a investigar, o bien observando con detenimiento las referencias de unos cuantos artículos relevantes al caso, por ejemplo, mirando un índice temático.

En 1964, Garfield presentó el *Science Citation Index* (SCI), catálogo de indexación por citas para los artículos científicos, un indicador que, en muy poco tiempo, revolucionaría las políticas de ciencia y tecnología. Garfield consideraba factible analizar las mil revistas científicas más prestigiosas y estudiar para cada disciplina la dinámica de citas. Más tarde, se publicarían por parte del Institute of Scientific Information (ISI),

³ Es interesante observar que su tesis doctoral en lingüística por la Universidad de Pennsylvania consistió en desarrollar un algoritmo para traducir nomenclaturas químicas y convertirlas en fórmulas moleculares.

con sede en Filadelfia, fundado y dirigido por Eugene Garfield, el *Social Sciences Citation Index* (SSCI) que recogería la dinámica de citas de las revistas correspondientes a las ciencias sociales, economía, sociología, política, etc., el *Arts & Humanities Citation Index* (AHCI) para las artes y las humanidades y, finalmente, el *Journal Citation Report* (JCR) cuyo objetivo sería presentar un indicador de las revistas científicas más prestigiosas, en cualquier área. Garfield llamó a este indicador Factor de Impacto (FI), y es actualmente uno de los indicadores más importantes para conceder becas, proyectos de investigación, sexenios, etc.

El FI, que ha logrado sobrevivir hasta hoy, pretende calcular la calidad de las revistas. Calcula el impacto o influencia que tiene una revista sobre la comunidad científica en un determinado año y lo mide tomando, para el año en curso, el cociente entre el total de las citas (de las revistas analizadas) que apuntan a la revista a lo largo de los dos años precedentes por el total de artículos publicados por esa misma revista en el mismo intervalo de tiempo. Por ejemplo, el FI de una revista en 2018 es el cociente entre la suma de citas a esa revista durante el período 2016-2017 y el total de artículos publicados en dicha revista en el período 2016-2017. Por tanto, un FI elevado muestra que dicha revista ha tenido una buena visibilidad en el mundo académico y sus artículos han influido en las investigaciones de un mayor número de colegas, lo cual ejerce una fuerte presión en los investigadores por publicar en dicha revista con el objetivo de difundir lo máximo posible sus trabajos.

$$\text{Factor Impacto 2018 (revista X)} = \frac{\text{Número de citas recibidas por la revista X en 2016 y 2017}}{\text{Número de artículos publicados por la revista X en 2016 y 2017}}$$

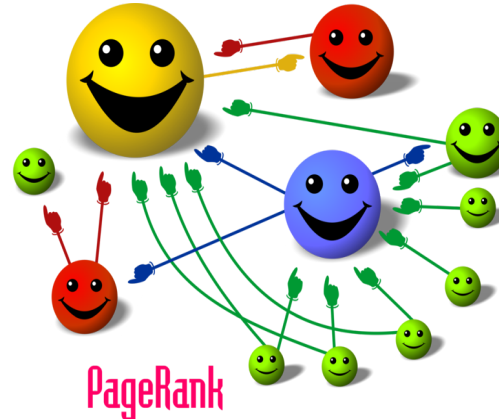
La intuición de Garfield fue la siguiente: la sección de referencias bibliográficas de un artículo académico muestra las fuentes sobre las que basa sus reflexiones y/o demostraciones el autor de dicho artículo; por consiguiente, cuando el autor cita esas fuentes las está destacando como valiosas para sus reflexiones y/o demostraciones. Y la inclusión de una fuente es una acción claramente intencional. Si un autor

considera que un artículo que ha leído no merece la pena, simplemente no lo citará; por el contrario, si lo considera valioso, lo citará. Siguiendo este razonamiento, podemos inferir entonces que el número de citas recibidas por una revista indica la calidad de la misma. Veamos un caso extremo: si el numerador del cociente del FI fuera una cifra extremadamente alta (es decir, que la revista recibe muchas citas) y su denominador fuera una cifra extremadamente baja (es decir, que la revista publica muy pocos artículos al año), significa que la revista de marras publica muy poco, pero aquello que publica es extremadamente bueno porque merece la atención de sus lectores. Estos lectores no se han limitado a leer sus contenidos, sino que han decidido citarlo, lo cual es una acción que podríamos llamar como de “agradecimiento intelectual”. Los investigadores y académicos, en su doble función de lectores y escritores de artículos, están diciendo con su cita que el artículo merece la pena.

Los fundadores de *Google*, Larry Page y Sergey Brin, no conocían la existencia de Shepard Citations, pero sí el mecanismo subyacente al FI porque ambos estaban realizando investigaciones (Page, Brin, Motwani y Winograd, 1999) y redactando artículos para obtener su doctorado en la Universidad de Stanford. Es así que se dieron cuenta que el modelo de Eugene Garfield de evaluar la calidad de los artículos mediante la indexación por citas, que a su vez estaba basado en la idea de indexación por citas de Franck Shepard, podía aplicarse también a la indexación de páginas web.

La lógica que subyace a *PageRank* es la misma que la que subyace al Factor de Impacto, si bien el primero es más sofisticado debido a la introducción de un criterio adicional, que aquí llamaremos *criterio de calidad*. Un vínculo de una página web A a una página web B es un “voto” de A a B. La importancia de una página depende del número de “votos” recibidos (*criterio de cantidad*), pero *Google*, además, analiza las páginas que emiten los votos, de modo que los votos emitidos por páginas que son en sí mismas “importantes” (porque reciben muchos votos) tienen más peso y ayudan a su vez a hacer que otras páginas sean importantes (*criterio de calidad*).

Figura 4. Imagen descargada de la web corporativa de Google.



Fuente: Imagen perteneciente al dominio público.

Podemos ver, desde la perspectiva de la teoría de grafos, que las páginas web hacen las veces de nodos y los vínculos o enlaces entre páginas web hacen las veces de aristas:

- **Criterio de cantidad** para medir la relevancia o calidad de un sitio web: sumar todos los vínculos o enlaces entrantes (*inlinks*) a ese sitio web. Ejemplo: si un sitio web A tiene 100 *inlinks* y un sitio web B tiene 50 *inlinks*, entonces A es considerado más relevante que B.
- **Criterio de calidad** para medir la relevancia o calidad de un sitio web: los vínculos o enlaces pueden tener pesos diferentes. Los enlaces provenientes de sitios web que tienen muchos *inlinks* valen más que los enlaces provenientes de sitios web que tienen pocos *inlinks*. Ejemplo: si un sitio web C tiene un *inlink* proveniente del sitio web A (100 *inlinks*) y un sitio web D tiene un *inlink* proveniente del sitio web B (50 *inlinks*), entonces C es considerado más relevante que D porque su *inlink* es más valioso.

Vemos pues que el algoritmo *PageRank* no cuenta el número de visitas que tiene un sitio web, porque las visitas no capturan una acción intencional de aprobación. Un usuario puede llegar a un sitio web (lo que cuenta como 1 visita) por diversas circunstancias, lo cual no significa que le guste ese sitio web una vez que está ahí. Sin embargo, crear un enlace a esa página (1 link) sí es una forma premeditada de valorar ese sitio web como relevante. Y vemos además que

PageRank otorga más peso a los enlaces (*links*) que vienen de sitios web relevantes, o sea, de sitios web que son de fiar (lo que aquí hemos llamado *criterio de calidad*, pero que muy bien podría llamarse también *criterio de autoridad*).

4. El funcionamiento del algoritmo de Facebook

El algoritmo de Facebook se llama *EdgeRank*⁴. En los últimos años *EdgeRank* ha sufrido modificaciones importantes, pero la filosofía de lo que realmente importa sigue siendo más o menos la misma. De todas formas, independientemente de los cambios que haya sufrido, nuestro interés aquí no es el mismo que el de los informáticos, los publicistas y los *community managers*, que sí tienen que conocer al detalle el algoritmo para maximizar sus ventas, así que nos valen unas pocas pinceladas de cómo está diseñado para captar la esencia del asunto.

Al acceder a su cuenta de Facebook, los usuarios pueden ver las actividades de sus amigos en un hilo de noticias llamado “muro”. En este muro, Facebook solo les muestra aproximadamente el 15% de las actividades de sus amigos en Facebook. Es decir, que si un usuario tiene 100 amigos y cada amigo realiza 10 acciones⁵ en un día, el número de acciones que verá en su muro a lo largo de ese será de más o menos 150, porque este es el 15% de 1000 acciones. Naturalmente, Facebook realiza este filtro porque considera, acertadamente, que el usuario no está interesado en ver las 1000 acciones, de modo que su algoritmo *EdgeRank* está diseñado para averiguar cuáles son los amigos cuyos mensajes probablemente interesarán más a dicho usuario.

A la hora de analizar las interacciones de un usuario con su muro, Facebook ha reducido todas las acciones posibles a sólo cuatro tipos básicos: hacer un clic, dar “me gusta”, hacer un comentario, compartir una noticia. Y cada una de estas acciones básicas tiene un peso:

- **Hacer clic** en la noticia o foto de un amigo es la acción que menos vale.
- **Dar “me gusta”** a una noticia o foto de un amigo significa que el usuario tiene un interés mayor en la noticia o foto que si solo hace clic.
- **Hacer un comentario** a una noticia o foto de un amigo significa que el usuario tiene aún más interés en la noticia o foto que si sólo da a “me gusta”.
- **Compartir la noticia** o foto de un amigo con el resto de amigos implica respaldarla aún más; uno puede poner en riesgo la propia reputación compartiendo cosas con otros, de modo que esta acción es la que más vale.

Cada acción se representa como una flecha (esto es, una arista con punta) que parte de un nodo (el usuario) y llega a otro nodo (la noticia o foto). Por ejemplo, un clic es la acción (representada por una arista) de pulsar con el ratón sobre la foto; dar a “me gusta” es una acción cualitativamente diferente que consiste en hacer clic sobre el icono de un dedo con el pulgar que está asociada a la misma noticia o foto. Y lo mismo para las acciones de comentar y compartir.

Y, como vemos, Facebook asigna a estas acciones pesos diferentes. Cuanta más intencionalidad pone el usuario a la acción, mayor peso asignará Facebook a dicha acción. A lo que aquí llamamos *intencionalidad* (Brentano, 1974; Dennett, 1991; Searle, 1992; Searle 1999), los expertos en comunicación y redes sociales lo llaman *engagement* (Bowden, 2009; Gambetti y Graffigna, 2010; Dessart, 2017) porque consideran con acierto que cada acción mide el compromiso del usuario con respecto a la acción realizada sobre el objeto digital de marras (una noticia, una foto, un video, etc.).

Figura 5. Imagen de elaboración propia.



Fuente: Karim Gherab. Reconocimiento-CompartirIguale (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

⁴ La palabra ‘edge’ en inglés significa arista, de modo que ‘edge rank’ significa un ranking de aristas. Queda clara pues la relación de *EdgeRank* con las Teoría de Grafos.

⁵ Estas acciones pueden ser de distinto tipo: subir una imagen o un video, decir cómo te sientes, responder a otro usuario, compartir una noticia subida por un amigo, hacer clic en una imagen o un video, dar “me gusta” al mensaje o la foto de alguien, y un largo etcétera.

Finalmente, como si de un cóctel se tratara, Facebook añade este ingrediente (el peso) a su coctelera (el algoritmo), mezclándolo con otros dos ingredientes: la afinidad (*affinity*) y el tiempo (*time decay*)⁶. El resultado final de todo esto es el siguiente: aquellos amigos que más acciones (cuantitativa y cualitativamente) reciben del usuario en un día específico son los que más probabilidades tienen de aparecer en el muro de ese usuario en los días siguientes.

Figura 6. Imagen de elaboración propia.



Fuente: Karim Gherab. Reconocimiento-CompartirIgual (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

Vamos a ver a continuación varios ejemplos de los que se conoce como tecnologías o sistemas de recomendación (*recommender systems*), que básicamente son algoritmos diseñados para sugerir a un cliente un producto o un servicio mientras navega por un sitio web. Por ejemplo, Amazon fue de las primeras compañías en utilizar estas tecnologías: los libros sugeridos suelen aparecer bajo el epígrafe de “Compras relacionadas” o “Los usuarios que compraron este libro también compraron este otro”.

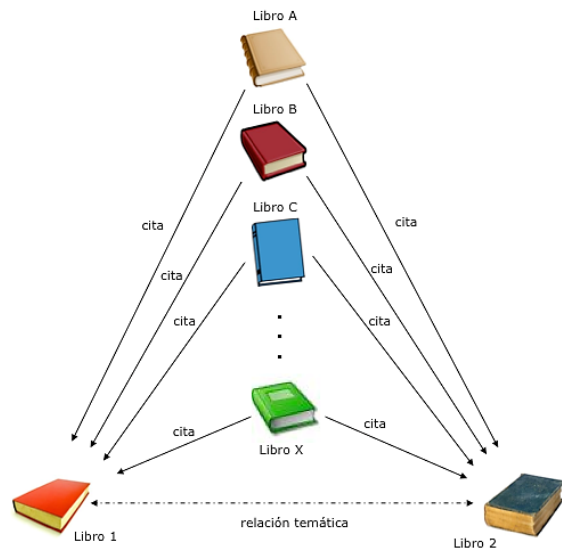
Los sistemas de recomendación no sólo sugieren productos o servicios, también pueden sugerir personas. Es habitual que redes sociales como LinkedIn, Twitter, Spotify y otras recomienden amigos con quienes conectar o a quienes seguir. En el siguiente apartado veremos la misma técnica aplicada a esquemas que asignan significados distintos a nodos y a aristas.

⁶ La afinidad (*affinity*) se refiere a aficiones comunes y otras interacciones, y el tiempo (*time decay*) se refiere a la obsolescencia de la noticia o foto, pero no entraré en el detalle de estos aspectos porque carecen de interés para nuestras consideraciones. El motivo es que en principio no tienen relación con disposiciones mentales tales como la intencionalidad o el *engagement*.

5. Una aproximación a los sistemas de recomendación

Como ya se ha dicho, no es este el lugar para explicar en detalle el funcionamiento de los sistemas de recomendación (Jannach, Zanker y Felfernig, 2010; Kembellec, Chartron y Saleh, 2014; Aggarwal, 2016), por lo que mostraremos sólo esquemas muy simplificados de algunos de estos mecanismos en aras de visualizar con más facilidad algunos conceptos subyacentes. El primer ejemplo consiste en mostrar las citas de unos libros a otros. Como puede inferirse rápidamente, si un número muy alto de libros (A, B, C, D, ..., X) citan dos libros específicos (aquí denotados como Libro 1 y Libro 2), entonces es muy probable que ambos libros traten temas similares. Fíjese que el algoritmo no necesita saber de qué tratan los Libros 1 y 2, y aún así, será capaz de detectar una relación entre 1 y 2. Y el algoritmo dará un mayor peso ponderado a la relación entre 1 y 2 cuanto mayor sea el número de libros que los citan.

Figura 7. Imagen de elaboración propia.

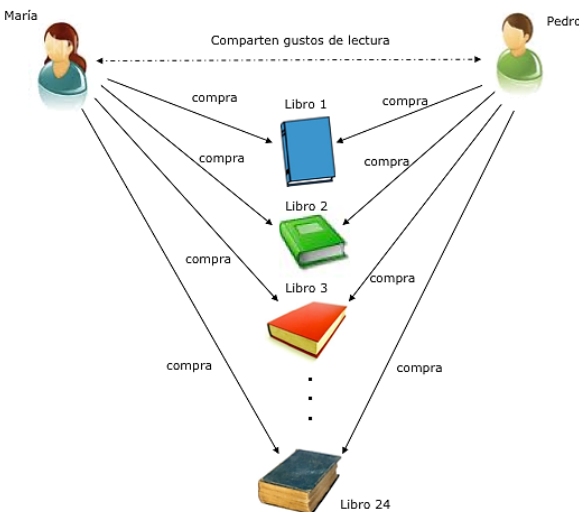


Fuente: Karim Gherab. Reconocimiento-CompartirIgual (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

En el gráfico siguiente, vemos que el algoritmo ha detectado que María y Pedro comparten gustos de lectura o tienen interés en temas similares. Esto es debido a que ambos, con

el tiempo, han ido comprando libros, y hasta en 24 ocasiones estas compras han sido coincidentes, es decir, María y Pedro han coincidido en comprar un mismo libro 24 veces. La conclusión es sencilla: si María y Pedro compran los mismos libros significa que les interesan los mismos temas. Por consiguiente, la siguiente vez que María compre un libro, el algoritmo recomendará ese libro a Pedro, y viceversa. Más aún, cuantas más compras coincidentes haya entre María y Pedro, mayor será el vínculo entre ambos en el sentido de que el algoritmo dará un mayor peso a esta relación (ver línea o arista discontinua en el gráfico siguiente).

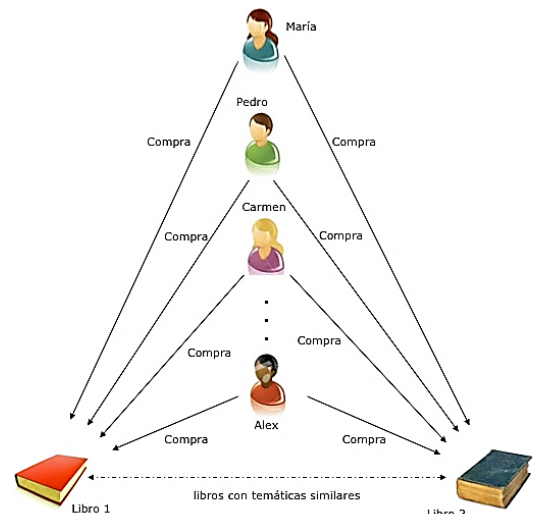
Figura 8. Imagen de elaboración propia.



Fuente: Karim Gherab. Reconocimiento-CompartirIgual (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

En el siguiente gráfico la relación temática que se extrae no es entre los compradores sino entre los libros. Si muchos usuarios de una librería online compran dos mismos libros (en el gráfico de abajo, los libros etiquetados como 1 y 2), entonces podemos sacar dos conclusiones: o bien se trata de dos *best sellers*, o bien ambos libros tratan muy probablemente del mismo tema. Supongamos que descartamos la primera opción; entonces cuantos más lectores comunes tengan ambos libros, mayor será la probabilidad de que ambos libros traten sobre lo mismo, es decir, tengan la misma temática.

Figura 9. Imagen de elaboración propia.



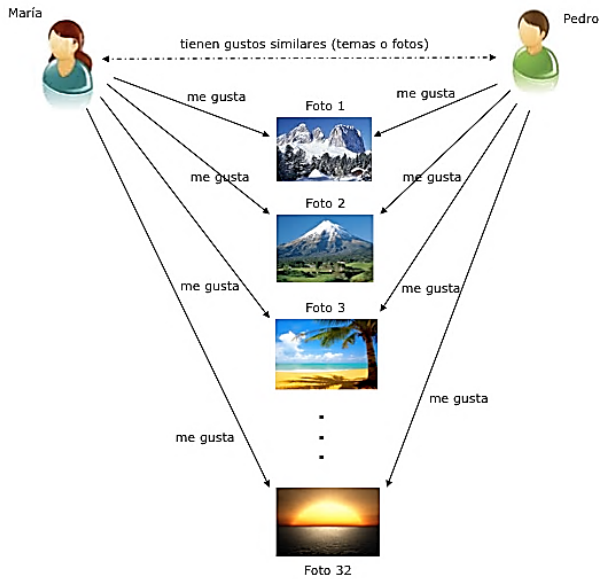
Fuente: Karim Gherab. Reconocimiento-CompartirIgual (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

En los gráficos anteriores hemos visto ejemplos de recomendaciones en las que las flechas (flechas unidireccionales de línea continua) partían de personas y apuntaban a libros. Aquí hemos ejemplificado con libros, pero los objetos a los que apuntan las flechas podrían ser perfectamente otras cosas, por ejemplo CDs, DVDs, fotografías, música, videos, etc.

Supongamos ahora que aplicamos la misma técnica a Facebook. Al igual que en los ejemplos anteriores con libros, las flechas parten de personas (sujetos) hacia objetos (aristas que unen nodos-sujetos a nodos-objetos), en este caso son fotos colgadas por otros miembros de esta red social. Lo interesante aquí es que las flechas no representan compras, sino otro tipo de acción. Concretamente, aquí la acción es pulsar el botón "me gusta", que Facebook y otras redes sociales similares ponen a nuestra disposición. Lo que consigue Facebook es que transformemos una información hasta entonces privada en una información pública. Y con toda esta información distribuida, ahora hecha pública mediante la acción de hacer clic en "me gusta", Facebook logra hacer explícito un conocimiento distribuido que tiene un inmenso valor publicitario y, por consiguiente, también monetario.

Veamos a continuación dos esquemas similares, pero con el triángulo invertido:

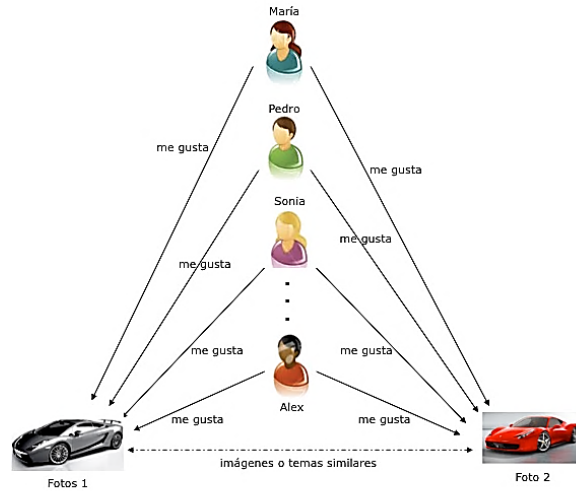
Figura 10. Imagen de elaboración propia.



Fuente: Karim Gherab. Reconocimiento-CompartirIgual (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

En el gráfico anterior, a María y a Pedro les gustan el mismo tipo de fotografías colgadas por sus respectivos amigos. En este caso, se trata de fotografías de paisajes. Cuantas más fotos de paisajes coincidan ambos en clicar “me gusta”, mayor será la probabilidad de que a María y a Juan les guste viajar o hacer, pongamos por caso, turismo rural. O simplemente les gustan las fotografías de paisajes más que las fotografías de ciudades. No es necesario insistir en la impagable información que Facebook puede extraer de estas acciones. Sin embargo, que Facebook capture esta información no es en sí mismo ni bueno ni malo. Será bueno o malo dependiendo de cada usuario. Si un usuario de Facebook desea que esta red social le ofrezca productos a medida o le muestre publicidad que puede interesarle, entonces el usuario saldrá ganando. Por ejemplo, al detectar que María y Pedro tienen gustos similares en cuanto a fotografías de paisajes, entonces podría recomendar a Pedro un curso al que se ha apuntado María. O incluso podría sugerirles una excursión a ambos con alguna agencia especializada en turismo de aventura.

Figura 11. Imagen de elaboración propia.



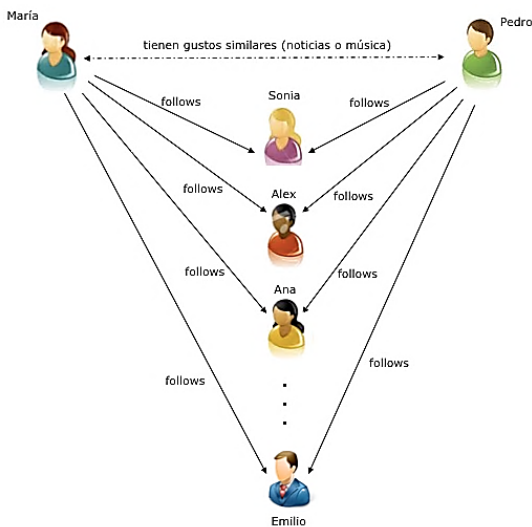
Karim Gherab. Reconocimiento-CompartirIgual (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

En el gráfico anterior la situación se invierte. Supongamos que unos usuarios de la red social han colgado unas fotos de coches (deportivos). Con el tiempo, a medida que las fotos se van compartiendo de unos usuarios a otros y van circulando de unas redes de amigos a otras redes de amigos diferentes, imaginemos que las fotos acaban llegando al muro de los cuatro usuarios que vemos en el gráfico (María, Pedro, Sonia, Alex), y de algunos otros (muchos o pocos). Hasta este momento, ningún empleado de Facebook ha visto estas fotos, ya que hay miles o cientos de miles de fotos que se cuelgan diariamente en la red social. En otras palabras, digamos que Facebook aún no sabe que el contenido de las fotos son imágenes de coches deportivos. Sin embargo, Facebook puede tener un algoritmo informático que detecta cuando varias personas hacen clic en las mismas fotos, e infiere que las fotos de marras probablemente tienen alguna similitud, por ejemplo, contenidos afines. Si se da este caso, cuantas más personas hagan clic en ambas fotos, mayor será la probabilidad de que las fotos 1 y 2 tengan algo en común. En este caso, se trataría de coches deportivos. Nótese que el algoritmo de Facebook sería capaz de recomendar a Alex una nueva foto que gustara a la vez a María, a Pedro y a Sonia (muy posiblemente se trataría de la imagen de un coche deportivo) sin que ningún empleado de Facebook haya intervenido

en el proceso. El algoritmo de Facebook simplemente estaría sacando ventaja de una serie de comportamientos colectivos, extrayendo y compartiendo un conocimiento tácito hecho explícito a través de clics de ratón. Imaginemos que se trata de una canción en lugar de una foto. Si el algoritmo no hubiera detectado y recomendado ese objeto (en este caso, una canción) a Alex, éste tal vez nunca hubiera sabido de la existencia de esta canción (quizá desconocida) que, al escucharla, descubre que le encanta.

Imaginemos a continuación cómo se aplicaría esta técnica en redes sociales como Twitter o Spotify, dos redes sociales con las que ejemplificaremos un caso diferente al mostrado en el caso anterior, en el que imaginamos un algoritmo que aplicara esta técnica en Facebook. Twitter trata de noticias, mensajes cortos que permiten muy pocos caracteres: los usuarios de Twitter comparten noticias o enlaces a velocidad de vértigo. Spotify es una red social especializada en música: se pueden compartir y recomendar canciones. En ambos casos se puede seguir (*follow*) la actividad de otro usuario, acaso porque pensamos que tenemos gustos comunes o porque nos interesan las noticias o las listas de canciones que ese usuario comparte, lee o escucha.

Figura 12. Imagen de elaboración propia.

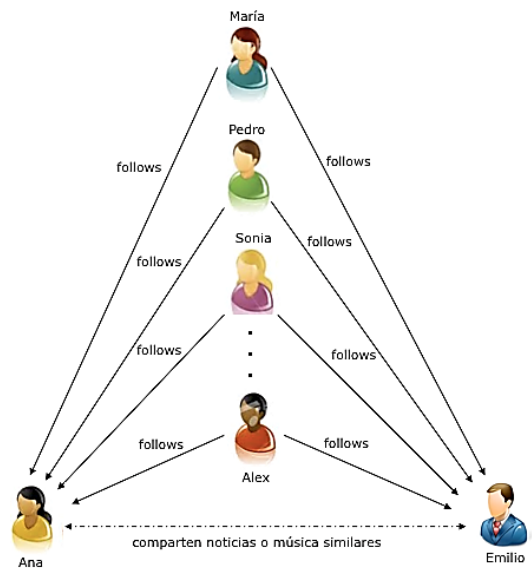


Fuente: Karim Gherab. Reconocimiento-CompartirIguual (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

Como puede verse en el gráfico anterior, en este caso, las flechas parten de unas personas y apuntan a otras personas (aristas que relacionan nodos-sujetos con otros nodos-sujetos), y representan un seguimiento de las acciones o actividades de estas últimas. Es fácil ver, siguiendo los ejemplos que hemos dado anteriormente, cómo Twitter o Spotify podrían hacer recomendaciones entre personas con gustos similares.

El gráfico siguiente es el mismo esquema, pero al revés. Si en el gráfico anterior se trataba de mostrar la relación de similitud e intereses comunes entre los seguidores (*followers*), en el siguiente destacamos la relación entre las personas que son seguidas, es decir, aquellas que publican información (noticias, canciones, etc.).

Figura 13. Imagen de elaboración propia.



Fuente: Karim Gherab. Reconocimiento-CompartirIguual (CC BY-SA). <http://creativecommons.org/licenses/by-sa/4.0/legalcode>

Hemos visto ejemplos de muchos tipos. Queda claro pues que los nodos y las aristas pueden representar cualquier cosa. A continuación, y como conclusión, reflexionaremos sobre cómo la combinación de estas técnicas con otras que ya están llevando a cabo las plataformas de *streaming*, tales como Netflix, conforman un marco de captura de estados mentales que los estudios de comunicación en redes sociales deberán analizar en profundidad.

6. Consideraciones finales

En los gráficos anteriores, las aristas han representado diversas acciones: citar, seguir (*follow*), dar “me gusta”, compartir, comprar, comentar, vincular/enlazar una página web, etc. Las posibilidades son virtualmente infinitas, así como infinitos son el número de clics que hacen los millones de usuarios de estas plataformas. Aprovechando este nuevo escenario, una nueva disciplina se abre paso en esta selva de datos: el *big data*. Netflix tiene actualmente más de 140 millones de suscriptores y su crecimiento se apoya en una estrategia de personalización de sus contenidos a los gustos (conscientes e inconscientes) de sus usuarios. Para ello, la compañía convocó en 2006 un concurso para premiar con un millón de dólares a la persona o grupo que maximizara la eficiencia de su algoritmo de recomendación *Cinematch* para aplicarlo a su negocio⁷. El primer párrafo de las bases de la convocatoria era el siguiente:

Netflix is all about connecting people to the movies they love. To help customers find those movies, we've developed our world-class movie recommendation system: *Cinematch*SM. Its job is to predict whether someone will enjoy a movie based on how much they liked or disliked other movies. We use those predictions to make personal movie recommendations based on each customer's unique tastes. And while *Cinematch* is doing pretty well, it can always be made better.

El ganador del concurso fue anunciado tres años más tarde⁸, el 26 de julio de 2009. Sin embargo, tres años después, el 6 de abril de 2012, Netflix publicó un comunicado de prensa⁹ señalando que, pesar de mejorar las búsquedas en un 10%, el algoritmo que ganó el concurso no sería implementado finalmente por diversos motivos. Aún así, Netflix seguía insistiendo en la importancia de las tecnologías de recomendación en un epígrafe titulado “Everything is a Recommendation”, donde escribía lo siguiente: “We have discovered through the years that

⁷ <https://www.netflixprize.com/>

⁸ https://www.netflixprize.com/community/topic_1537.html

⁹ <https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429>

there is tremendous value to our subscribers in incorporating recommendations to personalize as much of Netflix as possible”.

Al ser una plataforma de visionado de películas a través de *streaming*, Netflix tiene la capacidad de medir el *engagement* de sus usuarios simplemente monitorizando sus clics. Una manera fácil e inmediata de conocer el interés de sus clientes en la serie, por ejemplo *Black Mirror*, sería comparar el número de descargas o de accesos a la serie en la Temporada 3 con el de la Temporada 1, de modo que podrían saber si el número de visitas ha aumentado o ha descendido, y en qué porcentaje. Pero también podrían, y de hecho lo hacen, calcular el *engagement* de sus clientes mediante otros tipos de clics mucho más sutiles:

- Identificar los días en que los usuarios miran según qué contenidos. Por ejemplo, Netflix ha descubierto que sus clientes ven programas de televisión en días laborables y películas los fines de semana.
- Capturar las valoraciones cuantitativas que han hecho los clientes de las películas y combinarlas/compararlas con las pausas y los rebobinados detectados. Los usuarios de Netflix hacen aproximadamente 4 millones de valoraciones al día. Además, Netflix ha descubierto que el sistema de valoración basada en pulgares, hacia arriba o hacia abajo, es mejor que el sistema basado en 5 estrellas¹⁰.
- Comparar los datos, clics y valoraciones mencionados arriba con las búsquedas que realizan los usuarios. Los usuarios de Netflix hacen aproximadamente 3 millones de búsquedas al día.
- Monitorizar comportamientos de navegación en la web. Esto abre la puerta a averiguar si un cliente navega y hace clics en películas románticas y a comparar con sus patrones de clics en escenas de amor en otras películas.
- Capturar los clics o hacer capturas de pantalla cuando los usuarios hacen una pausa, rebobinan o aceleran con el fin de tener un conocimiento cualitativo de los

¹⁰ <https://media.netflix.com/en/company-blog/goodbye-stars-hello-thumbs>

gustos y psicología de sus usuarios con el fin de conocerles mejor y ofrecerles contenidos más personalizados en el futuro.

- Descubrir cuando un usuario hace clic en la pausa (*tasa de interrupción*) y poco después abandona la plataforma (*tasa de finalización*), y averiguar si vuelve o no vuelve (*fidelidad*). El objetivo es detectar los contenidos que no gustan a las diversas tipologías de clientes. Netflix ha descubierto que estas mediciones (inconscientes) pueden ser más fiables que las valoraciones (conscientes) que hacen los clientes de una película o serie.

Todas estas mediciones tienen como fin mejorar su algoritmo de recomendación. Netflix quiere ofrecer películas y series lo más ajustadas posible a los gustos y la psique del cliente para minimizar la probabilidad de que abandone la plataforma.

Naturalmente, todas estas métricas serán utilizadas no sólo para fines publicitarios y comerciales, sino que serán muy útiles para establecer métricas políticas, para conocer a los votantes. Nótese que todas las acciones que realizan los usuarios, y que se convierten en nodos y aristas a manos de los matemáticos, son conscientes. Sin embargo, con la ayuda del *big data*, será posible encontrar patrones inconscientes subyacentes (Mayer-Schönberger, V. y K. Cukier, 2013).

La revolución científica del siglo XVII se basó no sólo en las genialidades teóricas de científicos tales como Copérnico, Kepler, Descartes, Galileo y Newton, sino que fue condición *sine qua non* la práctica experimental. El nacimiento y mejora de los laboratorios en los siglos XVIII y XIX consolidó la física que hoy conocemos. Todo ello nos dio un conocimiento cada vez más exacto del mundo físico: se establecieron entonces las *leyes del movimiento* del mundo material. No cabe duda de que hoy nos encontramos en un momento muy similar a aquel, y que las redes sociales son los *nuevos laboratorios* donde las *grandes tecnológicas* hacen ya experimentos¹¹

que, con el tiempo, conducirán a unas nuevas *leyes del movimiento* de lo mental.

¹¹ *The New York Times*. "La empresa que explotó millones de datos de usuarios de Facebook", 20 marzo 2018, <https://www.nytimes.com/es/2018/03/20/cambridge-analytica-facebook/>.

Referencias

- Adair, W. C. (1955). "Citation Indexes for Scientific Literature", *American Documentation* 6, pp. 31-32.
- Aggarwal, C. C. (2016). *Recommender Systems: The Textbook*. Springer.
- Bowden, J. (2009). "The Process of Customer Engagement: A Conceptual Framework". *The Journal of Marketing Theory and Practice*, Vol. 17, no. 1, pp. 63-74. <https://doi.org/10.2753/MTP1069-6679170105>.
- Brentano, F. (1874). *Psychologie vom empirischen Standpunkt*. Ed. Oskar Kraus, 2 vols. Leipzig: Meiner. Disponible en <https://archive.org/details/psychologievome00kraugoog>.
- Brin, S & L. Page (1998). "The anatomy of a large-scale hypertextual Web search engine", Vol. 30, No. 1-7, April 1998, pp. 107-117, disponible en [https://doi.org/10.1016/S0169-7552\(98\)00110-X](https://doi.org/10.1016/S0169-7552(98)00110-X)
- Dennett, D. (1991). *La actitud intencional*. Barcelona: Gedisa.
- Dessart, L. (2017). "Social media engagement: a model of antecedents and relational outcomes". *Journal of Marketing Management*, Vol. 33, No. 5-6, pp. 375-399. <https://doi.org/10.1080/0267257X.2017.1302975>.
- Gambetti, R. C., y G. Graffigna (2010). "The Concept of Engagement: A Systematic Analysis of the Ongoing Marketing Debate". *International Journal of Market Research*, Vol. 52, No. 6, 801-826. <https://doi.org/10.2501/S147078531020166>.
- Garfield, E. (1955). "Citation Indexes for Science: a New Dimension in Documentation through Association of Ideas", *Science*, Vol. 122, No. 3159, July 15, pp. 103-111.
- Garfield, E. (1957): "Breaking the Subject Index Barrier: a Citation Index for Chemical Patents", *Journal of the Patent Office Society*, Vol. XXXIX, No. 8, August, pp. 583-595.
- Jannach, D., M. Zanker, y A. Felfernig (2010). *Recommender Systems: An Introduction*. Cambridge University Press.
- Kembellec, G., G. Chartron, y I. Saleh (eds.) (2014). *Recommender Systems*. John Wiley & Sons.
- Mayer-Schönberger, V. y K. Cukier (2013). *Big data. La revolución de los datos masivos*. Editorial Turner.
- O'Reilly, T. (2005). "What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software", disponible en <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>.
- Page, L., S. Brin, R. Motwani & T. Winograd (1999). "The PageRank Citation Ranking: Bringing Order to the Web". Technical Report. Stanford InfoLab, disponible en <http://ilpubs.stanford.edu:8090/422/>.
- Searle, J. R. (1992). *Intencionalidad*. Editorial Tecnos.
- Searle, J. R. (1999). *Intencionalidad en la filosofía de la mente*. Ediciones Altaya.