

The use of big data and data science to organise contents in digital social platforms, and its influence in socialization ^{1*}

El uso del *big data* y la ciencia de datos para la organización de contenidos en plataformas sociales digitales y su influencia en la socialización

RUBÉN JIMÉNEZ MAJIDÍ
Instituto para el Conocimiento,
la Gobernanza y el Desarrollo globales (ICGD)
Especialista en medios digitales
Director de la agencia Vahid Digital
ruben@vahid.es
ORCID: 0000-0002-2814-9697

Este artículo está sujeto a una: Licencia "Creative Commons Reconocimiento-No Comercial" (CC-BY-NC)

DOI: https://doi.org/10.24197/st.Extra_2.2021.140-155

RECIBIDO: 04/04/2021
ACEPTADO: 29/05/2021

Resumen: Las tecnologías de la información han multiplicado su capacidad de almacenamiento y de transmisión de información con el paso de los años permitiendo la concentración de enormes cantidades de información en vastas bases de datos. Este artículo hace un acercamiento a cómo una decena de plataformas web y medios de comunicación sociales hacen uso de la ciencia de datos para procesar e interpretar esta información, para analizar qué consecuencias tiene, por ejemplo, la selección y ordenación de contenidos, la predicción de hábitos de consumo o la influencia en los temas de conversación de los usuarios.

Palabras clave: comunicación; socialización; big data; ciencia de datos; redes sociales.

Abstract: Information technologies have increased their capacity to store and transfer information over the years, allowing large amounts of information to be concentrated in vast databases. This paper will take a look at how a dozen of the main web platforms and social media make use of data science to process and interpret this information, to analyse its consequences, for instance, the selection and organization of contents, the prediction of user consumption habits or the influence in users' topics of conversation.

Keywords: communication; socialization; big data; data science; social media.

^{1*} Este proyecto de investigación está vinculado a trabajos de investigación realizados por la agencia Vahid Digital.

1. INTRODUCCIÓN

Un individuo hace uso de las interacciones con sus semejantes para compartir —y generar— la experiencia humana. Este fenómeno, conocido como socialización, ha sido analizado por numerosos académicos desde distintos puntos de vista: por ejemplo, las implicaciones colectivas en la sociedad, o las implicaciones en el individuo y su relación con su círculo más cercano.

Este texto comienza explorando la definición de los medios de comunicación y haciendo un brevísimo repaso histórico de su desarrollo para identificar en cuáles de sus aspectos han tenido lugar los principales cambios con el paso de los años, con el fin de entender mejor el punto en el que se encuentran actualmente. A continuación, centrando la atención en dos aspectos característicos de los medios digitales modernos (la masificación del emisor y la instantaneidad) y que los diferencian de los demás medios tradicionales, analiza las implicaciones que el dominio sobre estas dos características tiene en la socialización de sus usuarios, es decir en la manera en los usuarios de los medios digitales construyen su visión del mundo y se relacionan con otras personas.

Las características de la masificación de los interlocutores y de la instantaneidad han empujado a las principales plataformas web que aglutinan las bases de datos de información más grandes que existen a invertir en el campo de la ciencia de datos con el fin de poner en orden toda esa cantidad de información para los usuarios.

1.1. Metodología

Para la elaboración de este artículo se ha analizado literatura existente y se ha hecho un estudio de casos de tipo descriptivo aplicado a varias plataformas web que hacen uso de la ciencia de datos. Las plataformas analizadas —Google, YouTube, Facebook, Instagram, Spotify, Netflix y Amazon— hacen uso de la ciencia de datos para seleccionar y ordenar contenidos, siendo ésta una de las fortalezas de sus servicios. Para analizar las plataformas se han identificado los tres modelos de negocio —no excluyentes— que emplean principalmente: suscripción, venta de producto y publicidad. También se ha observado qué aspectos del servicio que ofrecen requieren de ciencia de datos para ser viables. Generalmente, son aspectos relacionados con la personalización del aspecto de la plataforma para cada usuario ofreciendo, de la masa de contenidos, aquellos que puedan lograr que el usuario interactúe más con la plataforma.

Para analizar las plataformas se ha hecho una observación de su funcionamiento y sus funcionalidades, así como una lectura de sus políticas de uso y privacidad.

1.2. Una perspectiva coherente de las tecnologías de la comunicación

Tal como defiende Umberto Eco en su obra *Apocalípticos e integrados* (1964) cada vez que aparece un nuevo medio de comunicación surgen entre los estudiosos de la comunicación o la sociología dos tipos de pensamientos: uno que concibe al nuevo medio de manera crítica por las repercusiones negativas que implica a nivel social, «los apocalípticos»; y otro opuesto que realza sus aspectos positivos augurando que terminará reemplazando a otros medios anteriores, «los integrados». Finalmente, lo que según Eco siempre se ha demostrado es que ninguno de los extremos se cumple, sino que el nuevo medio desplaza ligeramente a los anteriores reubicando su espacio en la vida social, pero no necesariamente sustituyéndolos. La televisión desplazó a la radio del centro de los salones del hogar, pero nunca la sustituyó, al igual que la radio no reemplazó al periódico, ni internet está reemplazando a ninguno de ellos.

El presente texto no pretende ni presentar a los nuevos medios como problemáticos para la sociedad, ni tampoco ver todas las nuevas posibilidades que traen como ninguna suerte de salvación para la sociedad. Lejos de ello, la intención es comprender algunas de las consecuencias que ha tenido la emergencia de los nuevos medios de comunicación digitales en las masas de sus usuarios, y que cada una de estas tecnologías no es neutra sino que la propia decisión de utilizarlas lleva consigo algunos valores implícitos que aceptamos, frecuentemente de manera inconsciente.

2. LA SOCIALIZACIÓN A TRAVÉS DE LOS MEDIOS DE COMUNICACIÓN

Como se ha mencionado, podríamos dividir la socialización en dos perspectivas: la colectiva y la individual. El primer punto de vista (colectivo) lo analizaron Berger y Luckmann en *La construcción social de la realidad* (1966). Berger y Luckmann describen la comunicación como la manera en que los individuos o grupos intercambian información, y en base a la cual crean una visión de la realidad. Este entendimiento de que construimos socialmente la realidad a través de los medios implica que los medios de comunicación tienen un rol fundamental en dar forma a la sociedad, dado que nos ofrecen las pistas de aquello que no podríamos conocer de otra manera.

Por su parte, los escritos de Erving Goffman (1959) se adentran en el segundo punto de vista (individual) analizando la manera en que un individuo se presenta ante su círculo interpretando, ya sea consciente o inconscientemente, una determinada visión de sí mismo. Construye el «yo» que proyecta a través de su interacción con otros. Desde esta perspectiva más individual, los medios de comunicación sirven de herramientas para diseñar esa presentación.

Desde que existen, los medios de comunicación han jugado siempre un papel fundamental en la sociedad, pero la profundidad del impacto que tienen ha variado conforme los medios se han desarrollado. Un breve análisis histórico permite

examinar en qué aspectos han tenido lugar estos cambios y así comprender mejor el momento actual de los nuevos medios digitales. Entendemos por nuevos medios digitales aquellos que, habilitados por Internet, permiten una conexión instantánea a esta red global. Aunque históricamente ha sido sencillo identificar cada medio de comunicación, Internet ha alcanzado unos niveles de ubicuidad tan altos que en ocasiones se hace incluso complicado definir qué es, como medio de comunicación.

3. UN BREVE REPASO HISTÓRICO AL DESARROLLO DE LOS MEDIOS DE COMUNICACIÓN

Los medios de comunicación han cambiado mucho desde que, hace apenas unas décadas, C. Shannon y W. Weaver diseñaron su conocido modelo de comunicación (fig. 1) en su artículo «*Teoría matemática de la comunicación*» (1948). Sin embargo, la esencia de este modelo sigue intacta: la comunicación es la transmisión de paquetes de información desde un emisor hasta un receptor, a través de un canal, que interfiere mediante ruido y retardo. El desarrollo de las tecnologías de la comunicación ha llevado a otros modelos que describen de distintas maneras el proceso comunicativo, pero todos suelen compartir esta descripción básica.

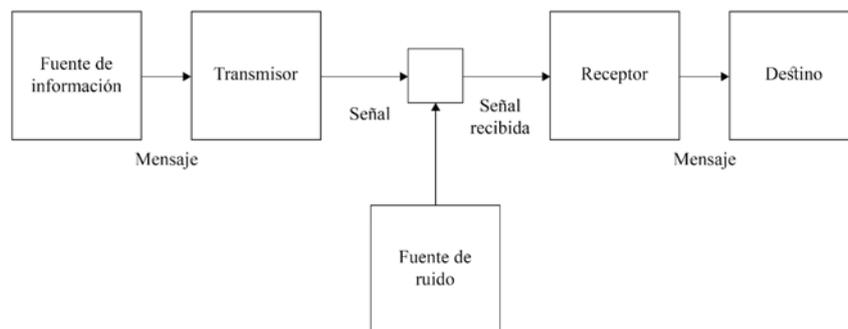


Fig. 1 Representación del modelo de comunicación de C. Shannon y W. Weaver. Elaboración propia.

3.1. Del telégrafo a las redes sociales

Uno de los principales aspectos en los que se han desarrollado las tecnologías de la comunicación es en la *masificación* de los distintos elementos del modelo de comunicación. Desde el mensaje sencillo y unidireccional, pero instantáneo, de un único emisor a un único receptor que habilitó la revolucionaria invención del telégrafo de Samuel Morse en 1844; hasta los vastos océanos de bytes de información multimedia entre los cuales navegan simultáneamente millares de emisores y millares de receptores en las redes de Internet.

Una posible manera de describir el desarrollo de las tecnologías de la comunicación a lo largo de la historia es verlo en los siguientes cuatro aspectos: la masificación del emisor, la masificación del receptor, la creciente velocidad de transmisión hasta rozar la inmediatez, y el alcance geográfico del mensaje (fig. 2). Podría haber otros aspectos que se podrían valorar, pero estos cuatro tal vez son los que describen de la mejor manera los cambios que han supuesto las distintas tecnologías como para comprender los nuevos medios.

Canal	Emisor	Receptor	Alcance						
			Limitado	Almacenable	Transportable	Larga distancia	Instantaneidad	Global	
Comunicación primitiva	Individual	Individual	Limitado al alcance del sonido y la vista						
Escritura (3500 AC)	Individual	Individual	Limitado al alcance de la vista						
Tambores, cuernos... (3000 AC)	Individual	Grupos pequeños	Limitado al alcance del sonido						
Papiro y tinta (3000 AC)	Individual	Individual	Movimiento físico						
Sistema postal (2400 AC)	Grupos pequeños	Grupos	Movimiento físico						
Telégrafo (1844)	Individual	Individual	Larga distancia. Instantáneo						
Teléfono (1876)	Individual	Individual	Larga distancia. Instantáneo						
Radio (1906)	Grupos pequeños	Masas	Larga distancia. Instantáneo						
Televisión (1923)	Grupos pequeños	Masas	Larga distancia. Instantáneo						
Internet (1969)	Masas	Masas	Global. Instantáneo.						
Telefonía móvil (1986)	Individual	Individual	Global. Instantáneo.						

Smartphones (años 2000)	Masas	Masas	Global. Instantáneo.						
----------------------------	-------	-------	----------------------	--	--	--	--	--	--

Fig. 2. Tecnologías de la comunicación a lo largo de la historia, en función de su potencial de alcance. Elaboración propia.

Uno de los académicos más destacados que han abordado estudios acerca de la historia de las tecnologías de la comunicación es Marshall McLuhan. En *La Galaxia Gutenberg* (1962) McLuhan describe la historia de la comunicación con el objeto de identificar la influencia de cada tipo de medio en la psique humana, dividiéndola en cuatro grandes etapas históricas en función de la naturaleza de los medios de la época: la tribu oral, la cultura manuscrita, la galaxia Gutenberg y la era electrónica. Su famosa frase «el medio es el mensaje» (McLuhan, 1964), que dio título a una de sus principales obras, hace referencia a la influencia que a su parecer tiene la tecnología de comunicación escogida —el medio— en el propio contenido o en la interpretación del mensaje que se transmite. Con ello como referencia divide la historia de las tecnologías de la comunicación en las cuatro etapas mencionadas.

Las conclusiones a las que llegó McLuhan hace ya más de sesenta años dibujan en el horizonte del futuro un mundo caracterizado por los medios electrónicos al que hace referencia como una *Aldea Global* (1962:31-32). Describiendo las tecnologías como «extensiones de las facultades humanas» defiende que a través de los medios de comunicación electrónicos «hemos extendido nuestro sistema nervioso central hasta abarcar todo el globo, aboliendo tiempo y espacio, al menos en cuanto a este planeta se refiere» (McLuhan, 1964:3). Esta afirmación, que fue recibida con escepticismo y crítica al ser enunciada en los años 1960, resulta sin embargo muy familiar en la actualidad gracias a la penetración de los medios digitales.

Al observar las distintas etapas descritas por McLuhan se puede observar que, en realidad, cada una viene caracterizada por una pequeña revolución en uno de los elementos de la comunicación identificados por Shannon y Weaver:

La tribu oral consta de toda la comunicación previa a la invención de la escritura. La naturaleza de la comunicación entonces era fugaz, no había posibilidad de conservar un mensaje para permitir que viaje en el tiempo o en el espacio, por lo que su alcance se limitaba a aquellos que estaban en ese lugar y momento para escucharlo o verlo.

La cultura manuscrita supone todo un nuevo paradigma para la comunicación, y de hecho para la historia humana al ser considerada la culminación de la prehistoria. La gran característica que se agrega a la comunicación es la capacidad del *almacenamiento* de la información, extendiendo así el *alcance* de la comunicación al permitir que un mensaje viaje en el tiempo y en el espacio gracias a la recogida de mensajes en papiros y libros manuscritos. Aunque poco se conserva, hoy en día podríamos ser receptores de mensajes emitidos hace 5.000 años, y gracias al papiro una carta podía recorrer miles de kilómetros de distancia.

La galaxia Gutenberg hace referencia a la revolución que supuso en la historia de la comunicación la invención de la imprenta moderna, atribuida a Johannes Gutenberg. El elemento que se ve transformado en este caso es el *receptor*. Si atrapar el conocimiento en un libro supuso un gran avance en la comunicación, tener la posibilidad de replicarlo tantas veces como se deseara abría la posibilidad de comunicación de un emisor hacia las *masas* de receptores.

La era electrónica, la última de las etapas descritas por McLuhan, da comienzo en el año 1844 con la invención del telégrafo de Samuel Morse. Este es el primero de una larga serie de medios de comunicación electrónicos que surgen a lo largo de un siglo. El elemento de la comunicación que es expuesto a un gran cambio es la *instantaneidad* en largas distancias, pero otros medios electrónicos posteriores como la radio o la televisión incidieron también en el *receptor* de la comunicación extendiendo su alcance hasta permitir que un emisor pueda hacer llegar un mensaje hasta masas de personas.

Los medios digitales en red han supuesto un cambio tan grande que está abierto el debate sobre si se trata de una nueva etapa en la historia de la comunicación. Los cambios de paradigma son difíciles de identificar, y solo el paso del tiempo puede mostrar el alcance que tienen, pero las tecnologías de los medios digitales lideradas por Internet han supuesto la masificación en el último de los elementos de la comunicación que aún no se había alterado: el *emisor*. Las redes sociales modernas aglutinan mensajes de multitud de emisores en pantallas de seis pulgadas. De esta manera se combinan los atributos de masificación del emisor, masificación del receptor, instantaneidad y alcance global.

4. BIG DATA Y CIENCIA DE DATOS

4.1. Generación y almacenamiento de datos

Hasta ahora el artículo ha tratado de describir el proceso histórico por el que han pasado medios de comunicación: desde transmisiones de comunicación sencillas y lineales, hasta complejos mecanismos que interconectan audiencias masivas en forma de redes. Las plataformas virtuales modernas pueden albergar a millones de personas que son al mismo tiempo emisores y receptores de un número indeterminado de contenidos textuales o audiovisuales. A continuación, se analizarán algunos datos que permiten visualizar, con tolerancia a cierto margen de error, la colosal cantidad de información que se está generando y almacenando a través de los nuevos medios digitales y el ritmo al que ha crecido en estas últimas dos décadas:

Un estudio de la Universidad de Berkeley, titulado *How much information?* (Lyman, Varian, et al., 2003) estimó que a lo largo del año 2002 se almacenaron aproximadamente 5 exabytes de información nueva entre los soportes de papel, película, magnético y óptico. Cinco años más tarde, a lo largo de 2007 se estima que se generaron 271 exabytes de información (Hardy & Williams, 2010) es decir 54

veces más. En 2018, once años después, se generaron 33.000 exabytes (Reinsel, Gantz & Rydning, 2018), unas 122 veces más que en 2007, y el mismo estudio de IDC prevé que en 2025 se produzcan 479 exabytes de información diariamente, es decir cerca de 175.000 a lo largo del año.

Analizando los datos generados desde **redes sociales, exclusivamente a través de smartphones**, se transmitieron unos 4'6 exabytes de información en 2020 según los estudios de *Ericsson Mobility Visualizer* (s.f.), lo cual es aproximadamente el doble que en 2018, y catorce veces en comparación con el año 2012. Las previsiones son que en algún momento de 2023 esta cantidad se vuelva a duplicar. Es decir, a lo largo de esta década se ha mantenido un ritmo de duplicación cada 2 o 3 años.

4.2. Poner orden en el caos

Con semejantes cifras de vértigo, es obvio que la gran mayoría de la información almacenada no es relevante, ni tan siquiera es posible de consumir para los receptores. Por poner un ejemplo, a lo largo de 2020, cada minuto se han subido a YouTube una media de 500 horas de contenidos, es decir unas 720.000 horas de contenidos nuevos diarios. Visualizar los vídeos subidos en tan solo 24 horas llevaría algo más de 82 años día y noche, que es casualmente la esperanza de vida de una persona en España.

Para las plataformas que manejan estas cantidades de información esto supone una serie de desafíos. A un nivel básico se enfrentan a cómo poner orden dentro de este caos de información, y ofrecer al usuario el contenido que le será más relevante —el gran debate es si se trata del contenido más relevante para el usuario, o más bien el que más interesa a la plataforma—. A otro nivel, estas plataformas están encontrando que de entre todo el ruido de datos les es posible, mediante complejos algoritmos y fórmulas, interpretar el pensamiento colectivo de la comunidad, dibujar distintos patrones y *personas*, e incluso predecir comportamientos futuros en base a toda la información analizada de personas con perfiles similares. Lo «único» que esto requiere por parte de la plataforma es ser capaz de leer y analizar esos exabytes de datos.

5. EL USO DE LA CIENCIA DE DATOS POR PARTE DE ALGUNAS PLATAFORMAS SOCIALES DIGITALES

Las principales plataformas que alcanzan la suficiente complejidad como para poder —o necesitar— hacer uso de ciencia de datos son propiedad de grandes empresas con objetivos comerciales. En este sentido, se podrían identificar tres modelos de monetización de los servicios, no necesariamente excluyentes entre sí: el modelo de uso abierto financiado por la **publicidad** (Google, YouTube, Facebook, Instagram...); el modelo de **suscripción** (Netflix, Spotify, HBO...); y el modelo de

venta de productos (Amazon, Ebay...). En todos ellos la acumulación y el análisis de la mayor cantidad posible de los datos cedidos voluntaria o involuntariamente por los usuarios son gran parte del valor de la empresa, por lo que el hecho de ceder datos de consumo ya es en sí mismo un pequeño pago tácito que realiza el consumidor.

En la medida en que la cesión de datos es tan valiosa para la plataforma, se hace igualmente relevante la retención del usuario. Una mayor retención implica más información acumulada, más anuncios visualizados, más productos vendidos y más fidelización. Debido a ello se emplea la ciencia de datos para retener al usuario el máximo tiempo posible, convirtiéndose de esta manera la ciencia de datos en herramienta y objetivo al mismo tiempo.

5.1. Los algoritmos de Google dan forma a la web

El uso de algoritmos para la ordenación de resultados tiene consecuencias importantes, particularmente en plataformas con la altísima tasa de mercado que tienen buscadores como Google o YouTube. Una de ellas es el surgimiento de toda una corriente —que se ha llegado a convertir en una profesión por sí misma— de ingeniería inversa para descifrar el funcionamiento del algoritmo y optimizar de esta manera el contenido para proyectarle una mayor visibilidad. Esta tarea es popularmente conocida como *posicionamiento S.E.O.*, por las siglas en inglés de *optimización para motores de búsqueda*. De esta manera la plataforma con buscador —cuyos usuarios dan por hecho que selecciona y ofrece los contenidos más relevantes de manera neutra— está indirectamente afectando al contenido mismo que generan los creadores de contenidos modifique su producción.

La influencia de los buscadores en los propios contenidos comenzó ocurriendo de manera esporádica en los inicios de estas plataformas, cuando sus algoritmos eran más sencillos. Pero conforme buscadores como Google han consolidado su posición de fuerza, han aprendido sobre el efecto que tienen y lo han empleado conscientemente. Google expone abiertamente (Google, s.f.) algunos de los principios que tiene en cuenta con el fin de promover entre los desarrolladores web que los tengan en cuenta. Mencionaremos tres ejemplos al respecto:

Uno de los aspectos a los que Google ha dado creciente importancia es al tiempo de carga de las webs: a menores tiempos de carga, mejor experiencia de los usuarios. Aunque a priori es un aspecto secundario a la relevancia del resultado, de esta manera Google busca ofrecer resultados de mejor calidad más allá del contenido de los mismos. Un segundo ejemplo se desprende del anuncio de Google en 2014 (Google Security Blog, 2014) de que las webs que tuvieran su certificado de seguridad (SSL) verían su presencia web beneficiada, y de manera más patente todavía cuando, años más adelante anunció que en las que no lo tuvieran Google Chrome mostraría al usuario una alerta advirtiéndole de que son inseguras. El anuncio de estas decisiones llevó a que, en cuestión de semanas, una oleada de sitios web implantaran certificados de seguridad en sus páginas. Un tercer ejemplo podría ser

cuando Google detectó que una importante parte del tráfico era móvil y sin embargo un porcentaje significativo de la web no estaba adaptada, y decidió anunciar abiertamente que iba a priorizar entre sus resultados a las webs que tenían bien adaptada la versión para móviles, con la consecuencia de que millones de desarrolladores web se lanzaron a adaptar los sitios web a la versión móvil para no perder su posición en Google.

5.2. YouTube, uno de los pioneros en la ciencia de datos

A pesar de estar especializada únicamente en vídeos, esta plataforma adquirida por Google en 2006 recibe desde hace unos años más búsquedas que los principales buscadores generalistas —solo por detrás del propio Google— como *Bing* o *Yahoo!* Para cada una de las búsquedas realizadas ordena entre los miles de millones de vídeos que alberga, y para cada vídeo visualizado hace uso de la ciencia de datos para ofrecer sugerencias de otros contenidos que puedan captar el interés del usuario y lograr que siga navegando en su red.

Aunque la necesidad de que la plataforma encuentre un modo de poner en orden la vasta cantidad de vídeos que alberga es una cuestión básica, las decisiones sobre qué criterios tener en cuenta y qué peso dar a cada criterio no están exentas de consecuencias sobre el producto que ofrece. Una de estas consecuencias es la mayor o menor visibilidad que obtienen unos u otros vídeos o creadores de contenidos.

Uno de los fenómenos a los que nos han habituado los medios digitales es el de lo *viral*, en lo que YouTube fue una de las primeras plataformas que gestó fenómenos globales en cuestión de días.

5.3. Almacenar e interpretar los datos que se ofrecen en las redes sociales

Cada una de las plataformas de redes sociales tiene su propia naturaleza y sus propias reglas, pero al mismo tiempo comparten unas lógicas subyacentes. Habitualmente, o al menos entre las que se conocen como las principales redes sociales —Facebook, Instagram, Twitter y la propia YouTube— su principal fuente de monetización es la venta de espacios publicitarios. Para maximizar estos ingresos las plataformas tienen como uno de sus objetivos retener el mayor tiempo posible a sus usuarios y fidelizarlos para que regresen con mayor frecuencia. Cuanto más tiempo pasan en la plataforma, más anuncios consumen.

Una de las claves de la publicidad tradicional siempre ha sido conocer el *target* al que se dirige. Desde las revistas a los programas de radio o televisión, la publicidad siempre ha sido orientada al tipo de público interesado en cada contenido. Las redes sociales cambian un poco las reglas del juego, pero hacen posible, más que nunca, definir perfiles concretos de usuarios, lo cual hace que sea mucho más jugoso para los anunciantes.

Estas plataformas han descubierto que pueden interpretar la información que van dejando sus usuarios en las redes para generar sus perfiles y conectarlos con anunciantes que busquen dichos perfiles. Por ejemplo, Facebook pregunta a un usuario cuándo se ha graduado de qué estudios, o cuándo ha tenido un hijo. Con esa información podrá vender los espacios publicitarios que visualiza a compañías que ofrezcan productos o servicios para profesionales de su área, o para padres de bebés, niños o adolescentes en función de cuántos años han pasado desde el nacimiento de su hijo.

Esto es un ejemplo sencillo, pero da una idea del interés que tienen plataformas como Facebook en recopilar tanta información como les sea posible sobre todos los aspectos posibles. Algunos datos los preguntan explícitamente, mientras que otros los deducen mediante interpretación: la temática de los grupos a los que se une un usuario o con los que interactúa, los intereses de su red de contactos, la temática de los vídeos o imágenes a los que reacciona, la duración del tiempo en el que se detiene sobre qué contenidos, etc. Otra fuente de información para vender espacios a los anunciantes son los temas de conversación entre los usuarios. Si se habla de la compra de un tipo de producto, la plataforma puede deducir que el usuario está interesado en invertir para ese producto y ofrecérselo.

Es, por lo tanto, muy importante para estas plataformas, por un lado, recabar tanta información como les sea posible sobre sus usuarios y, por otro lado, desarrollar la capacidad de interpretar dichos datos y generar los perfiles de sus usuarios. En este proceso, retener al usuario el máximo tiempo posible cumple dos propósitos: acumular más información sobre él —refinando al mismo su comprensión de perfiles similares— y poder venderle mayor número de anuncios.

A día de hoy no hay total claridad con respecto hasta qué punto *Facebook Ads* y *Google Ads* hacen lectura de palabras clave en las conversaciones privadas que tienen lugar en las plataformas de las que son propietarias sendas compañías matrices —Facebook, Instagram y WhatsApp en el caso de la primera, y YouTube, Google Maps, Gmail, etc. la segunda—. Facebook afirma no hacerlo con WhatsApp a raíz de la encriptación de extremo a extremo, y Google anunció hace años que dejó de sacar información de Gmail. La posibilidad técnica está ahí: es posible analizar los temas de conversación, los intereses y preocupaciones de las personas. Se trata de servicios gratuitos y exentos de publicidad que se alimentan únicamente de la información que les proporcionan los usuarios, para emplearlos en sus respectivas plataformas publicitarias

5.4. Las predicciones de Google Maps

Hace años que Google dejó de ser un buscador. Su entrada en el sector del *smartphone* a través del sistema operativo de código abierto *Android* que puede ser empleado por cualquier fabricante de móviles, les ha permitido entrar, en su sentido más literal, en los bolsillos de millones de personas. Esto permite a Google recabar

datos de millones de usuarios y predecir movimientos, ya no solo de sus intereses y búsquedas online, sino de sus movimientos geográficos por el planeta.

Su aplicación Google Maps, adaptada a la navegación por GPS, estima con apenas margen de error la duración de un trayecto en base a observaciones previas de otros usuarios, incluso prediciendo horas de congestión de tráfico o atascos. No muchos usuarios saben que pueden acceder a su historial de movimientos por el globo terráqueo de hace incluso 10 años.

Google es capaz de averiguar si un trayecto se ha realizado en coche, a pie o en bicicleta, cuánto tiempo se pasa en cada comercio o lugar de ocio, qué días de la semana o franjas horarias son las más concurridas en cada ubicación, etc. Al mismo tiempo, los empresarios complementan esa información directamente mediante el sistema *My Business*.

5.5. Netflix y Spotify, las nuevas industrias culturales

El modelo de comunicación de Netflix o de Spotify es ligeramente distinto al de las otras plataformas mencionadas en este artículo. Siguiendo los criterios descritos en el segundo epígrafe, no cuentan con un número indeterminado de emisores, sino que la propia plataforma es la única emisora —aunque se puede argumentar que comparte contenidos creados por terceros— sin embargo, a diferencia de la televisión o de la radio, no hay un único flujo de comunicación sino cada usuario accede a contenidos «bajo demanda» y bajo recomendación.

La propia Netflix (s.f.) explica en su web que ha «creado un complejo sistema de recomendaciones patentado» el cual describe someramente. Entre los aspectos sobre los cuales recopila y analiza datos se encuentran: el historial de contenidos consumidos, la valoración que les ha dado el usuario, la hora del día en que se consume el contenido, la cantidad de tiempo que se invierte en visualizar el contenido y el tipo de dispositivo que se está empleando. Toda esta información se cruza con la de otros millones de usuarios, y se correlaciona con los géneros, actores y años de producción de los contenidos para prever qué contenido será de interés del usuario:

Recopilamos comentarios de cada visita al servicio de Netflix y entrenamos de forma continua a nuestros algoritmos con esas señales para mejorar la precisión de su predicción sobre lo que es más probable que te guste ver. Nuestros datos, algoritmos y sistemas de computación siguen alimentándose unos a otros para producir recomendaciones actualizadas con el fin de proporcionarte un producto de tu agrado. (Netflix, s.f.)

A diferencia de las redes sociales, la propia Netflix es una de las principales creadoras del contenido que ofrece, con lo cual toda esa información recabada le permite detectar los elementos y tendencias necesarias para refinar los productos audiovisuales que más le interesa empaquetar para el usuario. Como resultado del

análisis de consumo de estos millones de visualizaciones Netflix es capaz de identificar, por ejemplo, que la duración idónea del capítulo de una serie para retener al usuario viéndolo de un tirón y estar aún dispuesto a ver otro capítulo más, es de unos 40 a 45 minutos. Tan seguro está Netflix de esto que le sale a cuenta no solo crear contenidos siguiendo los principios que identifica, sino también modificar productos audiovisuales ya finalizados para adaptarlos a sus criterios.

Este es el caso por ejemplo de la serie española *La casa de papel*, originalmente compuesta por nueve capítulos de unos 70 minutos de duración, que en su versión en otros idiomas cuenta con trece capítulos de unos 40 minutos de duración. Esto ha implicado incluso cambiar el orden de algunos acontecimientos del guion, o dejar algunos *cliffhangers* —un recurso narrativo que consiste en generar tensión creando suspense al final de un bloque del guion— a mitad de un capítulo.

Spotify es otro ejemplo similar, en este caso de la industria musical. Su amplia base de usuarios permite dibujar perfiles de personas con gustos similares y asociar gustos musicales. De una manera simplificada, si dos usuarios escuchan mucho una serie de canciones en su mayoría coincidentes, Spotify identificará aquellas similares en las que no coinciden y se las recomendará al otro en listas de reproducción como la denominada *Descubrimiento semanal*, en la cual Spotify ofrece 30 canciones que el usuario no ha escuchado nunca, basadas en el consumo de otros usuarios con quienes coincide en gustos.

5.6. Amazon, el rey de las *marketplace*

Esta empresa que comenzó como una librería online, aplica la ciencia de datos para las ventas de todo tipo de productos. Amazon fue una de las primeras grandes compañías en utilizar estas técnicas. Analizando las elecciones de compras de sus millones de usuarios Amazon realiza predicciones de ventas, lo que le permite hacer envíos a sus almacenes antes de que siquiera hayan sido demandados. De esta manera refuerza su argumento estrella de envíos al hogar en un solo día. Por supuesto, al igual que las otras plataformas, su complejo sistema de recomendaciones le permite no solo seleccionar productos en los que pueda estar interesado el comprador, sino orientar sus compras hacia los productos similares sobre los que probablemente no había pensado, pero en los que podría estar interesado. En este proceso, a pesar de ser una *marketplace* en la que diversos vendedores pueden ofrecer sus productos, son los productos propios de Amazon —particularmente los que tiene disponibles en su almacén— los que tienden a tener mayor visibilidad y presencia.

A diferencia de los ejemplos mencionados anteriormente, Amazon es una plataforma orientada a la venta de productos, por lo que se pone en entredicho la objetividad a la hora de ordenar los resultados que, si bien no está del todo clara, se presupone a las otras plataformas. El beneficio de la plataforma no es el mismo con la venta de unos productos u otros.

De los tres modelos de monetización descritos al comienzo de este epígrafe, Amazon ha sido capaz de combinar las tres: los usuarios pagan una suscripción anual por acceder al servicio (suscripción), pagan por los productos consumidos (ventas), y por el trayecto consumen publicidad pagada por otros proveedores (publicidad).

6. CONCLUSIONES

Este texto ha comenzado revisando los cambios que han experimentado los medios de comunicación a lo largo de la historia, tomando como referencia los elementos básicos de la comunicación: el emisor, el receptor, el tiempo de retraso, el alcance y el almacenamiento del mensaje. Este análisis nos ha llevado identificar que los momentos más importantes en la historia del desarrollo de las tecnologías de la comunicación han coincidido con una revolución en uno de estos aspectos.

Siguiendo la lógica detrás de la descripción de la historia de la comunicación desarrollada por McLuhan (1962) está abierta la cuestión de si nos encontramos o no frente a una nueva etapa posterior a las que describió, protagonizada en este caso por los medios digitales. Declarar si se trata de este caso o no escapa a las intenciones y posibilidades de este artículo, pero los argumentos expuestos indican que sí podría tratarse de una nueva etapa. La última de estas revoluciones, que ha dado lugar a los nuevos medios de comunicación digitales aglutinados por la ubicuidad de Internet, ha estado concentrada en la masificación de los emisores. Los grandes números de emisores, combinados con la instantaneidad, han producido a su vez también la masificación de la cantidad de mensajes, generando cantidades incalculables de datos. Una tendencia que, además, sigue en aumento.

La capacidad de procesamiento de dichos datos está mostrando ser tan compleja como valiosa. Tanto es así que el propio modelo de negocio de algunos productos digitales está basado en la sola recopilación de datos de los usuarios a cambio de servicios gratuitos. La ciencia de datos es la capacidad de dar sentido y analizar estas cantidades de datos. Gracias al refinamiento de estas técnicas, algunas grandes empresas están logrando crear perfiles muy afinados de sus usuarios, no solo identificando sus necesidades sino incluso adelantándose a ellos gracias a los patrones existentes que han aprendido a leer y predecir.

En una primera instancia, el surgimiento de las plataformas sociales parecía democratizar la comunicación poniendo a disposición de todos los usuarios la participación en estos nuevos medios de comunicación masivos. No obstante, con el paso de unos pocos años, esta masificación de contenidos junto a la búsqueda de maximización de beneficios y del tiempo de retención de los usuarios ha llevado a las plataformas a buscar fórmulas para filtrar y ordenar los contenidos. El solo hecho de no tener un criterio sencillo y claro para ordenar los contenidos —como podrían ser el orden cronológico de publicaciones, o el orden alfabético de productos— y más aún el filtrarlos, atenta contra la neutralidad de estos medios.

Un ejemplo de cómo el uso de la ciencia de datos está afectando a algunos aspectos de nuestra socialización como personas se puede observar al aplicar la teoría de la espiral silencio descrita por Noelle-Neumann (1977), adaptada a los medios digitales. Los usuarios en plataformas sociales como Facebook o Twitter tienden a agruparse para estar en contacto con aquéllos hacia quienes sienten mayor simpatía o hacia cuya visión sienten mayor sintonía. Al consumir los mensajes que recibirán en su red de contactos podrían sentir reafirmadas sus opiniones, o incluso considerarlas opiniones generalizadas al ver que tienen una representación en su entorno mediático.

Otro ejemplo de impacto en los procesos de socialización es la manera en que plataformas como YouTube o Instagram premian aquellos contenidos que logran mayor retención de usuarios otorgándoles aún más visibilidad, produciendo que los contenidos y emisores más populares tiendan a serlo todavía más, y los menos populares tener menos visibilidad.

Que los medios de comunicación de masas no son ventanas abiertas y objetivas a la realidad ya ha sido ampliamente estudiado en el pasado. Los algoritmos sobre la ciencia de datos, tal como describe este artículo muestran algunas de las limitaciones que las plataformas sociales tienen con respecto a su neutralidad.

REFERENCIAS BIBLIOGRÁFICAS

Berger, P. & Luckmann, T. (1966) *The social construction of reality: A Treatise in the Sociology of Knowledge*. New York: Anchor Books

Ericsson Mobility Visualizer: mobility data app (s.f.). Retrieved from:
<https://www.ericsson.com/en/mobility-report/mobility-visualizer?f=9&ft=2&r=1&t=12,13,14&s=4&u=3&y=2011,2025&c=1>
[accessed 1 December 2020]

Goffman, E. (1959) *La presentación de la persona en la vida cotidiana*. Amorrortu, Buenos Aires, 1993.

Google (s.f.) *Cómo funciona la búsqueda de Google: empoderar a los propietarios de sitios web*. Retrieved from:
<https://www.google.com/search/howsearchworks/mission/creators/> [accessed 1 December 2020]

- Google Security Blog (2014) *HTTPS as a ranking signal*. Retrieved from: https://security.googleblog.com/2014/08/https-as-ranking-signal_6.html [accessed 1 December 2020]
- Hardy, C. & Williams, S. (2010) Managing information risks and protecting information assets in a web 2.0 era, in *Implications for the Individual, Enterprises and Society*. Retrieved from: [https://domino.fov.uni-mb.si/proceedings.nsf/Proceedings/D706DFE93A1ACCF0C1257757003B55D7/\\$File/17_Hardy.pdf](https://domino.fov.uni-mb.si/proceedings.nsf/Proceedings/D706DFE93A1ACCF0C1257757003B55D7/$File/17_Hardy.pdf) [accessed 1 December 2020]
- Lyman, P., Varian, H. R., Swearingen, K., Charles, P., Good, N., Jordan, L. L., Pal, J., (2003) *How much information? 2003*, School of Information Management and Systems, University of California Berkeley, Berkeley, California, USA 2003. Retrieved from: <http://www.sims.berkeley.edu/research/projects/how-much-info-2003/> [accessed 1 December 2020]
- Netflix (s.f) *Cómo funciona el sistema de recomendaciones de Netflix*. Retrieved from: <https://help.netflix.com/es-es/node/100639> [accessed 1 December 2020]
- Noelle-Neumann, E., (1977) *La espiral del silencio: Opinión pública: nuestra piel social*
- Reinsel, D., Gantz, J. & Rydning, J. (2018) *The digitalization of the world, from edge to core*. An IDC White Paper – #US44413318. Retrieved from: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> [accessed 1 December 2020]
- Shannon, C. & Weaver, W. (1948) 'A mathematical Theory of Communication' *Bell System Technical Journal*, vol. 27 July 1948, p. 379-423
- McLuhan, M. (1962) *The Gutenberg Galaxy, the making of typographic man*. London: Routledge & Kegan Paul
- McLuhan, M. (1964) *Understanding Media: The extensions of man*. London: Routledge & Kegan Paul Ltd.