

## **La opinión pública entre la razón y el control social. Una actualización en la era del Big Data**

Carlos de Angelis\*

### **Resumen**

Los debates sobre cómo definir la opinión pública han generado innumerables discusiones a lo largo del siglo XX. En este trabajo se discutirán dos principales aproximaciones, la primera que define el fenómeno como el resultado de la racionalidad de los sujetos que contribuye al proceso de formación de la opinión y de toma de decisiones en una democracia; y la segunda como una modalidad de control social, cuyo papel sería promover la integración social y garantizar la obtención de consenso para basar las acciones y decisiones gubernamentales. En el presente trabajo se analizarán estas posturas centrales sobre la opinión pública para pasar a observar dos teorías recientes que retoman la tradición del control social: la teoría del espiral del silencio y la de *agenda setting* o imposición de agenda. Estos debates influyeron también en las concepciones del espacio público y en las formas de investigar el fenómeno, en este sentido se abordará la cuestión del estudio de la opinión pública, en particular la cuestión de los sondeos de opinión bajo el supuesto de la existencia de una esfera pública abordable metodológicamente y sus problemas. Luego se debatirán los cambios sociales recientes vinculados a las tecnologías info-comunicacionales y el surgimiento de una posible esfera pública virtual, enfocada especialmente en redes sociales como Facebook y Twitter, para finalizar se realizarán algunas aproximaciones para su estudio desde los conceptos de *Big Data* y *Data Mining*.

Opinión Pública - Control social – Big Data

---

\* Carlos de Angelis es Sociólogo y docente en la Facultad de Ciencias Sociales de la Universidad de Buenos Aires. Coordinador del Centro de Estudios de Opinión Pública de la Facultad de Ciencias Sociales – UBA.

## 1- Introducción

Dos miradas centrales se desarrollaron sobre la opinión pública a lo largo del siglo XX. Una posición planteaba que la opinión pública se constituía como una esfera racional de debate y de resolución de conflictos construyendo “verdades” socialmente legitimadas por mayorías. La otra postura en cambio identificaba a la opinión pública como un dispositivo de control social, ya sea como fuerza cohesiva que evitara la disolución de la sociedad, o como fuerza externa homogeneizadora de las posturas políticas con la finalidad que los sujetos de autorregulen.

A lo largo del presente trabajo se analizarán con detalle estas posturas para pasar a observar dos teorías recientes que retoman la tradición de la opinión pública como control social: la teoría del espiral del silencio y la de *agenda setting* o imposición de agenda. Luego se pasará a analizar un fenómeno hijo predilecto de la tradición racionalista: el estudio científico de la opinión pública, principalmente a través de los sondeos de opinión, bajo el supuesto de la existencia de una esfera pública abordable metodológicamente. Finalmente se debatirá la existencia de una esfera pública virtual a partir de la irrupción de Internet a partir de mediados de la década de los noventa, con especial énfasis en la emergencia de las redes sociales ya en la primera década del siglo XXI y algunas aproximaciones para su estudio desde los conceptos de *Big Data* y *Data Mining*.

### La opinión pública entre la racionalidad y el control social

Harwood Child en su libro *An Introduction to Public Opinion* (1949) reuniendo más de cincuenta definiciones de opinión pública buscando resaltar las dificultades de encontrar una definición consensuada para el concepto. Años después Noëlle Neuman llevó adelante el análisis de la definiciones reunidas por Child, llegando a la conclusión que “proceden de sólo dos conceptos diferentes de opinión pública” (2003, pág. 280). En efecto casi todas las definiciones surgirían de dos ideas de opinión pública:

- como resultado de la racionalidad de los sujetos que contribuye al proceso de formación de la opinión y de toma de decisiones en una democracia;

- como control social, cuyo papel sería promover la integración social y garantizar la obtención de consenso para basar las acciones y decisiones gubernamentales.

La existencia de la opinión pública como mecanismo de resolución racional que permitiera obtener “ciertas verdades” en el terreno de los hechos políticos y sociales supone la existencia de una esfera de hechos generados mediante “el libre debate público de individuos dotados de capacidad racionante y discursiva” (Germani, 1995, pág. 100), concepción propia de la tradición iluminista. Pero “estas verdades”, para Habermas (2009) no se producirán en el vacío, por el contrario tienen un sujeto: la burguesía y un objetivo: controlar la esfera estatal. La burguesía en el momento que se construye como sociedad civil se despliega de la sociedad estatal, organiza a la opinión pública como una forma de limitar la autonomía de los gobiernos y a la burocracia, donde ésta última tendría un papel creciente en las decisiones gubernamentales. Dos elementos esenciales para la invención de la opinión pública fueron el nacimiento de la prensa política y el surgimiento de espacios de discusión donde se reunía la burguesía europea donde se plantearía “una tendencia hacia la discusión permanente entre personas privadas, de ahí que dispusieran de una serie de criterios institucionales comunes” (2009, pág. 73), allí puede identificarse el germen de dos futuras instancias de control y articulación entre el Estado y la sociedad: el lobby empresarial y las Organizaciones No Gubernamentales (ONG).

Si bien la idea del control social como fuerza integradora de la sociedad ocupa un lugar relevante en la teoría social de raíz positivista desde los primeros trabajos de Herbert Spencer, se transforma en un concepto central de la sociología en las primeras décadas del siglo XX frente “a los efectos desintegradores del orden social que provocaban la expansión del capitalismo industrial y el desarrollo del imperialismo” (Olmo Oliver, 2005, pág. 74).

Una de las conceptualizaciones iniciales de control social fue desarrollada por Edward A. Ross y seguida por un conjunto de investigadores conocida como la “Escuela de Chicago” (Azpúrua Gruber, 2005). En términos generales los estudios sobre el control social apuntaban a los mecanismos sociales que mantienen el orden social en un sentido amplio, instituciones y prácticas que dotan a la sociedad de capacidades de regularse a sí

misma, acorde con los principios y valores deseados, es decir mediante la persuasión. Ross populariza el término “control social” en su obra *Social control, a survey of the foundations of order* (1901). Ross sostiene que el orden social en las sociedades avanzadas solo es posible mediante la solidaridad, y la cohesión social que se obtiene a través del control, entendido como “aquellos procesos sociales que coordinan las funciones del individuo y del grupo de manera organizada” (Berganza Conde, 2000, pág. 157). Se debe considerar que Janowitz plantea que la concepción de control social tiene un matiz progresista, pues es la alternativa a un control coercitivo, donde la organización de una sociedad descansa o se apoya predominante y esencialmente en la fuerza, en la amenaza y el uso de la fuerza (1975).

Robert Park fue quien plantearía que la opinión pública es una forma de control social. Este autor vinculado a la Escuela de Chicago desarrolló sus estudios bajo la doble influencia de la filosofía pragmatista norteamericana y el pensamiento idealista con foco en Simmel. (Nocera, 2008). Estas filiaciones teóricas condujeron a sus estudios sobre el papel de la comunicación y las prácticas interaccionales y su aporte sobre el concepto de distancia social. Park percibe en que las grandes urbes ocurren transformaciones desde una sociedad basada en relaciones primarias a otra de las relaciones secundarias (Park, 1925), donde el control social permite mantener la cohesión y la solidaridad dentro de la sociedad.

Opinión pública, las leyes y el ceremonial, son las formas específicas en las que el control social encuentra su expresión de una forma universal para Park (Berganza Conde, 2000). Aquí es relevante observar que para el autor la opinión pública sería una función latente, oculta, incluso para los propios sujetos. En este sentido para explicar los mecanismos a nivel psicológicos que emplean los sujetos para adaptarse a las demandas sociales Park retoma los conceptos de Gabriel Tarde, en cuanto a la imitación y sugestión. La categoría de imitación era central para Tarde puesto que “la cohesión social es resultado de esas leyes de imitación que operan a varios niveles, pero siempre consisten en subordinar los momentos racionales y creativos a otros más bajos y no creativos” (Laclau, 2005, pág. 61). Park se aparta de la mirada individualista de la opinión pública que luego se

impondría, pues considera que se impone a los sujetos como una realidad evidente y exterior, porque “la opinión pública no es una opinión que incumbe por igual a cada miembro del público, sino que se trata, más bien, “de una opinión o de un comportamiento que se presenta como ajeno frente a cada individuo y que se considera objetivo” (1996, pág. 400).

Park observa que los grandes periódicos, de surgimiento reciente, es el gran medio de comunicación de la ciudad. La opinión pública se apoya en la información suministrada por este medio, adelantándose por cuatro décadas a la teoría de la “Agenda Setting”. Sorprendentemente Park indica que el periódico reemplaza al chisme popular (Park, 1925) También había observado Gabriel Tarde la importancia de los medios en las nueva subjetividad de los individuos en las grandes ciudades, sugiriendo “que el lector no tiene conciencia de sufrir esta influencia persuasiva, casi irresistible del periódico que lee habitualmente” (2011, pág. 201).

## **Dos teorías sobre la formación de la opinión pública**

Dos de las teorías más importantes sobre la formación de la opinión pública se plantean en la década del setenta y tienen relevancia en la actualidad. Tanto “la espiral del silencio” y “el establecimiento de agenda” (*agenda setting*), se desarrollan sobre idea de la existencia de control social. Mientras que en la primera teoría la formación de la opinión pública se desarrollaría principalmente en forma endógena, en la segunda sería en cambio planteada en términos exógenos.

En la espiral del silencio se basa en la idea de que para evitar el aislamiento, un individuo puede evitar su propio juicio, como precio a pagar para estar integrado socialmente. (Noëlle-Neuman, 1998). En esta concepción el contexto mediato e inmediato ejerce presión sobre los sujetos para homogeneizar visiones y pareceres. Este proceso posee una dinámica propia, ya que los individuos se forman una idea del reparto y del éxito de sus opiniones dentro de su entorno social (el órgano “cuasi-estadístico); y esa lectura incentiva a que los sujetos expongan públicamente sus opiniones, en la medida que la consideren exitosas seguirán siendo expresadas. Sin embargo en cuanto los sujetos

comiencen a evaluar que las opiniones son débiles (encuentran resistencias), se inicia el proceso de cambio en el clima de opinión; y comenzarán a evitar a exponerlas, creciendo el temor a quedar aislados y con ello los sujetos preferirán el silencio al aislamiento, tratando de establecer los nuevos contextos de opiniones sin riesgos de sanciones. Esta idea ya había sido desarrollada por Goffman (2012) en los conceptos de “actuación” y “fachada”. Donde “actuación” es la actividad que tiene un individuo durante un período tiempo en co-presencia frente a una cantidad de espectadores y que posee cierta influencia sobre ellos y “fachada” es la dotación expresiva “empleada intencional o inconscientemente por el individuo en su actuación” (2012, pág. 36).

A principios de la década de los '70 McCombs y Shaw presentan su trabajo *The Agenda-Setting Function of Mass Media* (1972) donde desarrollarían el concepto del establecimiento de agenda para expresar la capacidad de los medios de comunicación para direccionar la atención de la opinión pública hacia temáticas particulares, priorizando ciertas cuestiones planteadas como sobresalientes y problemáticas para cada momento. Desde allí se desarrolla la idea de que la agenda de los medios no decide qué tiene que pensar el público, pero sí sobre qué tiene que opinar, determinando mapas cognitivos que guían a los sujetos, dicho de otra forma, los medios muestran “las “imágenes” de las noticias a modo de “ventanas” que presentan visiones limitadas del mundo exterior” (Rubio Ferreres, 2009, pág. 11). Esto puede suceder porque “los medios de masas representan todos los días el mundo como algo nuevo guiados por el código informable/no informable” (Luhmann, 2009, pág. 323).

No obstante para el análisis de los efectos de la agenda mediática por sobre la agenda pública se deben tener en cuenta elemento propios del *hábitus* de los individuos: la experiencia personal, género, el nivel educativo, el interés por la política y con respecto a la propia noticia vinculado al grado de exposición a las noticias y los efectos personales que le genera (McCombs, 2006). Debe tenerse en cuenta la tipología posterior que desarrolla McCombs en cuanto a una variedad de agendas o subagendas y la “combinación de agendas” o *agenda melding*, (2006, pág. 269) donde deben incluirse las agendas generadas por las redes sociales personales pero también las mediadas por medios electrónicos.

Más allá de medios y mecanismos involucrados, la teoría de la *agenda setting* establece un elemento esencial: la sociedad puede ser influenciada. La teoría de la espiral del silencio por su parte pone en juego un elemento irracional de los sujetos: el temor aislamiento.

### **Estudio “científico de la opinión pública” y predominio de las encuestas de opinión.**

La idea de que la opinión pública puede ser “medida” está presente desde las primeras décadas del siglo XX con los trabajos de los psicólogos y psicólogos sociales como Floyd Allport y Louis Leon Thurstone. El primero en su trabajo “Toward a Science of Public Opinion” de 1937, critica los usos ficcionales del concepto de opinión pública en la utilización de conceptos no científicos, tal como una “alma colectiva”. En cuanto a la personificación Allport sostenía que el tratamiento dado hasta el momento a la noción era de “una ficción relacionada a la aplicación de la noción de ser colectivo super-orgánico” (pág. 8), que en definitiva dejan de lado al individuo y sus decisiones. Por otra parte refuerza esta idea cuestionando uso de tres falacias: la grupal, la de la inclusión parcial y la falacia periodística. También cuestiona la *teoría emergente*, la *teoría eulogística* y la *ficción ideacional* de la opinión, concepciones que diluyen desde su punto de vista los procesos individuales en favor de generalizaciones colectivas. Frente a las dificultades para identificar los múltiples elementos que intervienen en la formación de la opinión pública Allport propone observar los fenómenos de la opinión pública en base un programa que muestra que son esencialmente “instancias de comportamientos” individuales.

En el marco de cierto predominio del modelo de las ciencias naturales, donde las ciencias sociales debería explicar fenómenos formulando leyes generales (Wilson, 1990) dos “tecnologías aplicadas contribuyeron en la imposición de las encuestas como estrategias racionales para la investigación de la opinión pública. Por un lado, el muestreo estadístico y por el otro la estandarización de los cuestionarios con la introducción de las escalas. También los principios estadísticos expresaron una traducción de la democracia a la investigación social en “la equivalencia de las unidades nombradas. Estadísticamente cada individuo vale como otro” (Reynié, 2012, pág. 42).

La “muestra” como una selección aleatoria de sujetos que pertenecen a una población puede ser una idea hoy intuitiva pero significó un avance decisivo en las estrategias cuantitativas de investigación social. Se debe recordar que en 1936 (en los años previos a la Segunda Guerra Mundial), George Gallup mediante una encuesta realizada con una muestra de cinco mil casos distribuidos por Estados Unidos “acertó” el triunfo del demócrata Roosevelt, por sobre el republicano y “aislacionista” Alf Landon cuyo triunfo era asegurado el sondeo de la revista *The Literary Digest* realizado mediante el conteo de las preferencias electorales a vuelta de correo de la masiva entrega de cupones a sus suscriptores. En este sentido, Gallup diría en 1939 que la muestra le permitiría “sondear la opinión de un electorado miniaturizado” (2012, pág. 58), mostrando que el método también poseía un elemento central de las ciencias naturales: podía predecir.

Por otra parte, las escalas de medición contribuyeron a proveer un marco estandarizado para la construcción de indicadores. Evaluar propiedades que pertenecen a estados mentales, como actitudes, valores, sentimientos y opiniones en general, sobre las cuales no existe un acceso directo y que tampoco poseen unidades de medida (Marradi, Archenti, & Piovani, 2010), resultaba un desafío que fue encarado por un grupo de psicólogos como Louis Leon Thurstone (Thurstone, 1931); Rensis Likert (1932) y Louis Guttman (1949) entre otros que buscaron la configuración de diversas escalas que permitieran capturar estados internos de los sujetos, a través de sus declaraciones a través de diferentes arquitecturas de los cuestionarios.

A lo largo del siglo XX bajo la representación de imparcialidad y cientificismo las técnicas de encuestas mediante muestreo se impusieron como fórmula dominante para la investigación en opinión pública, con especial predominio en períodos electorales, sin embargo, a partir de los años setenta los sondeos y en particular las empresas que suelen presentar este tipo de estudios, fueron blanco de diversas críticas, como por ejemplo que suelen entrevistar a personas que no tienen opinión formada sobre los temas planteados (Champagne, 2005); que generan de prácticas performativas a través de la difusión de datos de determinadas temáticas (Garrigou, 2007); y que simplifican la realidad a través de recortes arbitrarios en la difusión de algunos datos (Wolton, 2008).



### **“Escuchar las redes: ¿es escuchar a la opinión pública?: El ascenso del *Big data*”**

La idea que Internet recrea las condiciones de funcionamiento de una “esfera pública de característica virtual” (Papacharissi, 2002) en términos “habermasianos” resulta estimulante en términos de las posibilidades que abre para su investigación. Esta imagen de Internet reproduciendo parte o todo el tejido social, se basa por un lado en el acceso rápido y barato a la red, y en su capacidad para reunir personas superando fronteras geográficas e incluso idiomáticas.

Se debe destacar que las implicaciones de la revolución digital en la esfera pública vinculadas a las nuevas modalidades de participación tiene sus defensores y detractores, pero bajo la consideración generalizada que se está frente a un fenómeno inédito de potencialidades inexploradas (Gracia Pecourt, 2015). Entre sus potencialidades las tecnologías de la información tendrían un rol relevante en la eliminación del déficit de información, incrementando la transparencia del sistema político (Van Dijk, 2000) y lo que es fundamental la interactividad de los usuarios limitarían el acciones de las elites políticas posibilitando la creación de un gobierno abierto y responsable ante la sociedad. Estas acciones generarían nuevas modalidades o variantes de la democracia donde las tecnologías de comunicación: tele-democracia, ciberdemocracia y democracia electrónica (Harto de Vera, 2006).

Para los partidarios de la tele-democracia la democracia representativa tradicional está en crisis puesto que no responde al incremento de la pluralidad y sus nuevas demandas, causando apatía, frustración y alineación del electorado. Nuevas formas democracia directa serían posibles mediante el uso de las tecnologías de la información. En una dirección similar, los partidarios de la democratización electrónica buscar perfeccionar el sistema representativo con énfasis en incrementar y ampliar los flujos de información hacia la ciudadanía, buscando que todo tipo de información pase del dominio privado al público. Sostienen que modalidades de voto electrónico (como por ejemplo [change.org](http://change.org)<sup>†</sup>) para decisiones puntuales, permitirían fortalecer las relaciones entre representantes y

---

<sup>†</sup> Plataforma mundial de peticiones públicas

representados multiplicando los debates en torno a temáticas relevantes. Finalmente, los ciberdemócratas dan un paso más allá planteando ya no la reparación de la democracia tradicional, sino el establecimiento de una “verdadera” cuyo poder y fortaleza dependerían de sí misma y por otra parte la búsqueda del bienestar material vinculado a las nuevas formas de producción (y distribución) de riquezas. Esta perspectiva se vincula a cierta visión que visualiza al Estado como una amenaza y al rechazo a ciertas jerarquías que el poder estatal genera, la posibilidad de establecer relaciones horizontales, permitiría recuperar valores comunitarios.

Más allá de estas perspectivas la interactividad propia de la Web 2.0 habría comenzado por disputar algunas plazas a los actores tradicionales que organizaban al menos discursivamente la formación de la opinión pública como periodistas, encuestadores y políticos potenciando a dimensiones de una esfera pública en proceso de redefinición y que tiene un carácter transnacional y “glocalizado” (Sampedro Blanco & Resina de la Fuente, 2010), lógica particular que surge de combina (o directamente mezclar) lo global y lo local como es por ejemplo, el uso de neologismos anglosajones.

Sin embargo, también hay cuestionamientos a las posturas que consideran que Internet favorece la generación de una nueva y renovada esfera pública de ciudadanos informados, por su parte Polat (2005) cuestiona el automatismo de suponer que Internet conduce a una sociedad más informada. La información disponible en la red necesita ser procesada por el usuario con el fin de tener un significado, sin embargo el usuario tiene la capacidad de interpretar sólo una cierta cantidad de información, y esta capacidad está estrechamente vinculada a su capital cultural, y educativo. Por otra parte, los motores de búsqueda operarían sistemáticamente para favorecer a algunos sitios web y si bien Internet ofrece fuentes alternativas de información, la mayoría de las personas traspasan sus hábitos del mundo físico a Internet, donde finalmente los principales medios de comunicación dominan el entorno en línea noticias políticas. También se debe considerar que incluso más allá de la universalización del acceso a Internet lo más probable es que las desigualdades en el mundo real continúen en el virtual donde quienes más se beneficien sean los que ya cuentan con una ventaja en términos de ingresos, las capacidades y el acceso a los vínculos políticos, los cuales son importantes predictores de la participación.

Otras críticas insisten que Internet tendría efecto despolitizante además de la tendencia a generar información con sesgos (Hall, 2012) por la proliferación del entretenimiento basura; además se debe considerar la censura creciente sobre los flujos de información; la multiplicación de contenido xenófobo, violento y de propaganda pro-gubernamental; y la vigilancia de los contenidos publicados: la facilidad para establecer identidades por parte de empresas y gobierno (Morozov, 2006), donde incluso las empresas de tecnologías tendría la capacidad de generar un Estado paralelo que “no pretende mejorar el bienestar de los ciudadanos, sino aprovecharse de sus actividades para incrementar el bienestar de las grandes empresas”.

Más allá de la discusión sobre la capacidad de Internet para generar una esfera pública virtual existe un consenso en que gran parte de la vida social se refleja en los flujos informativos de internet. Esto genera un nuevo desafío en términos de investigación sobre como acceder y procesar al flujo incesante de información. “Big Data” es el nombre que toma este mundo virtual de información y Data Mining o Minería de datos, emerge como la nueva estrategia para la investigación de la opinión pública. Si bien se pueden tomar diversas definiciones, “Big Data” se refiere en general a la práctica de combinar grandes volúmenes de datos de diverso origen, y “Data Mining” al análisis de los mismos mediante la utilización de sofisticados algoritmos para interpretarlos (European Data Protection Supervisor, 2015). Este conjunto de técnicas de análisis de datos ha ido cobrando protagonismo especialmente. No existe aún una denominación específica en castellano para este conjunto de procedimientos, cuya traducción más aproximada sería minería de datos. Existen muchas definiciones para ésta nueva “tecnología”. Una de las más inclusivas expresa que *data mining* es la ciencia de extracción de información útil de grandes grupos de datos, una nueva disciplina que aprovecha la ventaja de la intersección de la estadística; las máquinas de aprendizaje; el aprovechamiento y manejo de base de datos; el reconocimiento de patrones; la inteligencia artificial y otras áreas (Hand, Mannila, & Smyth, 2001). Por el uso de las estadísticas a la que la definición hace referencia se refiere preferentemente a técnicas multivariadas, como análisis factorial; clúster o árboles de decisión, entre otras. El concepto de máquina de aprendizaje, o “*learning machine*” en inglés, se refiere a los programas preparados para encontrar patrones entre grandes

volúmenes de datos, encontrando modelos implícitos y realizar predicciones basadas en una información parcial o incompleta. El desarrollo de la inteligencia artificial (AI) está implícita en la máquina de aprendizaje, y se refiere al desarrollo de algoritmos específico para el tratamiento de los datos y término de *Data Warehouse* o “almacén de datos”, muestra un aspecto central estas técnicas que se enfoca a que el descubrimiento de patrones dentro de un grupo de datos establecidos “naturalmente”.

En sus inicios tanto Big Data como Data Mining se orientaron a dar respuesta al mundo de los negocios como forma de explorar y conocer los datos acumulados en enormes bases de datos, por ejemplo, datos de compras en supermercados, registros de uso de tarjetas de crédito, y las estadísticas gubernamentales. De hecho uno de las primeras técnicas de análisis se llamó precisamente “*Market basket analysis*”, es decir el análisis de la cesta de compras del supermercado, buscando asociaciones significativas en los datos de compra del cliente. La solución típica llevó a la generación de la clásica regla de “las personas que compran pañales son propensos a comprar cerveza ” (Raeder & Chawla, 2010). Sin embargo, velozmente las técnicas fueron expandiéndose hacia el análisis de otros tipos de datos como imágenes de los cuerpos astronómicos, bases de datos moleculares, y registros médicos.

El rápido desarrollo del “*Big Data*” se basa en un factor obvio: el crecimiento exponencial la producción y el almacenamiento incesantes de grandes volúmenes de datos, se comienza a acompañar de la capacidad para analizar, comprender y aprovechar el valor de esos datos. Estas relativamente nuevas modalidades de análisis también alcanzaron a las ciencias sociales, donde esos masivos flujos de datos, los científicos sociales se enfrentan a nuevas fuentes para acceder e interpretar y particularmente con vinculado a fuentes de datos provenientes de Internet. Como expresa David Lazar:

*“Vivimos la vida en la red. Cuando nos despertamos a la mañana comprobamos nuestro correo electrónico, hacemos una llamada telefónica rápida, caminamos al aire libre (nuestros movimientos son captados por una cámara de vídeo de alta definición), tomamos el autobús (deslizamos nuestras tarjetas de transporte) o en coche (utilizando una tarjeta a través de los peajes). Llegamos en el aeropuerto, compramos*

*un sándwich con una tarjeta de crédito antes de subir al avión, y comprobamos nuestros teléfonos poco antes del despegue. La visita al médico o al mecánico de automóviles, generan registros digitales de nuestros problemas médicos o del automóvil. Escribimos en nuestro blog confiando al mundo nuestros pensamientos y sentimientos, mantenemos los perfiles de redes sociales personales que revelan nuestras amistades y nuestros gustos...”* (Lazar, 2009, pág. 721)‡.

El puñado de actividades planteadas en el párrafo van dejando “rastros digitales” que analizados globalmente permiten el acceso a la vida de los individuos y con el potencial evidente de permitir una nueva mirada de la sociedad con una cantidad de información disponible inédita en la historia de la humanidad. En este sentido se “*Data Mining*” ha mutado a “*Web Mining*” o minería de la red, aplicando las técnicas de minería de datos a documentos y servicios del Web (Kosala & Blockeel, 2000). A partir de las huellas de quienes visitan un sitio en Internet dejan (direcciones de IP, navegador, etc.) los servidores almacenan en forma automática una bitácora de accesos. Las herramientas de la minería de la red analizan y procesan éstas bitácoras para producir información significativa. Los contenidos de Internet presentan varios tipos de datos, como texto, imagen, vídeo, metadatos se utiliza el término *Multimedia Data Mining* (minería de datos multimedia) como una instancia del *Web Mining* para tratar ese tipo de datos.

Los investigadores en temas económicos fueron de los primeros en utilizar estas herramientas para realizar seguimientos macroeconómicos y establecer tendencias (Choi & Varian, 2012), utilizando indicadores tales como los precios de vivienda, ventas minorista, inflación, empleo, comportamiento del consumidor y la previsión del consumo privado, donde se demuestra que la incorporación de datos provenientes de herramientas como Google Trends pueden mejorar las predicciones obtenidas mediante encuestas tradicionales (Vosen & Schmidt, 2013).

Debe tenerse en cuenta que en términos metodológicos la técnica de investigación de minería de datos establece diferencias sustanciales con las modalidades tradicionales de registro de datos cuantitativos, en especial con los sondeos de opinión. Ya no hay muestreo

---

‡ La traducción es nuestra

en el sentido tradicional del término, ni una intervención directa del investigador en campo donde a través de los cuestionarios y encuestadores interrogaba en forma directa a las unidades de análisis seleccionadas. Tampoco están disponibles las hipótesis tradicionales donde el investigador “aislaba” variables independientes y variables dependientes. Aquí los datos están disponibles en forma simultánea, por ese motivo se habla de “exploración de los datos”, y es posible vincular estos procedimientos a las técnicas de observación de fenómenos. Por esto también van creciendo dos estrategias adicionales frente a los resultados obtenidos: la visualización de los datos (y resultados), y el “*Text Mining*” o minería de textos. La exploración visual de datos permite observar en forma sencilla patrones observables en forma intuitiva y no requiere la comprensión de algoritmos matemáticos o estadísticos complejos (Keim, 2002). La minería de textos reconoce que gran parte de la información disponible en la Web está disponible en forma de textos, noticias de los periódicos, blog, entradas en redes sociales, etc., allí el *Text Mining* busca la extracción de información desde diversas fuentes escritas (Tan, 1999).

Las redes sociales, en especial Facebook y Twitter son dos ámbitos particulares donde la investigación desde las ciencias sociales y para los estudios de opinión pública en particular, han cobrado especial interés, dado que se han transformado en espacios concentrados de diálogo e influencia política.

Para Westling (2007) por ejemplo, Facebook se ha transformado en el mejor medio para lograr una verdadera esfera pública, donde millones de personas en todo el mundo lo han convertido en una arena para la comunicación. Facebook permite, por ejemplo que miembros de una comunidad, sea ciudad o pueblo aporten sus opiniones sobre un tema mientras que les da la flexibilidad de decidir cuándo y cómo contribuyen a la conversación, también permiten a los políticos comunicarse con los miembros de la comunidad, sin capacidad de imponer sus mensajes a cualquier persona. La contrapartida es que la plataforma se ha transformado en un espacio de campaña permanente (Larsson, 2016) por el uso masivo de estas red por parte del sistema político, desde políticos instalados que cuentan con gran estructura para la difusión de sus mensajes, hasta los que están desarrollando sus primeras acciones Esta suerte de “invasión” puede llevar a que muchos

usuarios se retiren o buscan redes alternativas. También debe tenerse en cuenta que es inusual que Facebook funcione como un generador de noticias, por el contrario la gran mayoría de la información proviene de los medios tradicionales, lo que hace que en la práctica Facebook funcione como complemento de las comunidades existentes en el mundo real, los problemas políticos del mundo real, y las noticias del mundo real.

Por su parte Twitter es caracterizado como una red social de microblogging por sólo permitir un número reducido de caracteres (restricción que hasta el presente Facebook no posee). La configuración predeterminada para los tweets (así son denominados los breves mensajes) es público, lo que permite a los usuarios a seguir a los demás y leer los tweets sin dar permiso mutuo (Jansen, Zhang, & Chowdury, 2009). Cada usuario tiene su página de Twitter donde las actualizaciones de los mensajes de los perfiles seguidos se agregan en una sola lista que permiten visualizar los mensajes en forma sencilla. Desde sus inicios en 2006 Twitter ha sido muy utilizada para obtener información sobre noticias, campañas políticas y recientemente difusión de productos comerciales. Algunos estudios sugieren que los usuarios de Twitter se adoptan más información de los líderes de opinión (es decir, celebridades, bloggers famosos, etc.), que de otros usuarios de redes sociales facilitada por la posibilidad de reenvío (Stieglitz, 2012). Por otra parte la presencia de un hashtags facilita las incorporaciones a comunidades virtuales temáticas. Una hashtag es una palabra cuyo prefijo de un Símbolo # que, añadido se puede añadir a un tweet a fin de indicar la temática sobre la cual se opina. Investigaciones recientes indican que el análisis de los tweets en un momentos determinado pueden ser predictores electorales (Tumasjan, A y otros, 2010), así como también pueden ser eficientes observadores del clima de opinión determinado frente a un acontecimiento como la muerte de una personalidad relevante (Calvo, 2015). No obstante, siguiendo a Calvo se debe identificar algunos obstáculos en el análisis de la *twittósfera* como por ejemplo la multiplicación de los llamados *fakes* “quienes anónimamente modelan y redirigen a la opinión pública en las redes” (2015, pág. 111) y los *trolls*, cuentas de identidades falsas que pueden “llevar adelante una multiplicidad de estrategias políticas que de otra forma sería imposibles” (2015, pág. 108).

Finalmente las estrategias más recientes de investigación apuntan a identificar los sentimientos de los sujetos que intervienen en la web. Estos caminos, que de alguna forma

buscan continuar a los psicólogos que en los años '30 del siglo XX crearon las escalas, cuentan con nuevas herramientas que apuntan a la identificación automática de los contenidos para la clasificación de mediante complejos algoritmos. El análisis de los sentimientos, ya tiene su denominación como "*Opinión Mining*" o minería de opiniones, y busca identificar las valoraciones, actitudes y emociones de la gente hacia entidades tales como productos, servicios, organizaciones, personas, problemas, eventos, temas, etc. (Liu, 2012). El análisis de sentimientos se centra principalmente en las opiniones que expresan o implican sentimientos positivos o negativos. Para posibilitar la definición será pues identificar el lenguaje natural como expresión de un sentimiento, que son subjetivos y socialmente compartido, por lo tanto para esto es imprescindible examinar una colección de opiniones de muchos sujetos en lugar de sólo una única opinión de una persona, para esto se deben construir un repertorio de opiniones y expresiones de sentimientos, y esta información es la se incorpora al algoritmo. Uno de los más empleados es el llamado API<sup>5</sup> que es precisamente un "clasificador de sentimientos" que es un método de aprendizaje automático para detectar la polaridad de los sentimientos específicamente en los tweets. Para cada tweet, la API convierte el contenido en una puntuación que indica si el tweet es positivo, negativo o neutro (Hoang, T. y otros, 2013). Uno de los desafíos de las técnicas es identificar los contextos de generación de los mensajes, así como también determinar los textos de contenido irónico que suelen transmitir un mensaje inverso al expresado.

---

<sup>5</sup> <http://help.sentiment140.com/api>



## Bibliografía

- Allport, F. H. (1937 (2002)). Toward a Science of Public Opinion. *Public Opinion Quarterly*, 1 (1), 7-23.
- Azpúrua Gruber, F. (2005). La Escuela de Chicago. Sus aportes para la investigación en ciencias sociales. *Sapiens. Revista Universitaria de Investigación*, 25-35.
- Berganza Conde, M. R. (2000). *Comunicación, opinión pública y prensa en la sociología de Robert E. Park*. Madrid: Centro de Investigaciones Sociológicas.
- Calvo, E. (2015). *Anatomía política de Twitter en Argentina*. Buenos Aires: Capital Intelectual.
- Champagne, P. (2005). Hacer hablar a la gente. El uso social de las encuestas en democracia. En L. Wacquant, *El misterio del ministerio. Pierre Bourdieu y la política democrática* (págs. 111-133). Barcelona: Gedisa.
- Child, H. (1949). *An Introduction to Public Opinion*. New York: John Wiley & Son.
- Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88 ( Issue Supplement), 2-9.
- European Data Protection Supervisor. (2015). *Meeting the challenges of big data. A call for transparency, user control, data protection by design and accountability*. Bruselas: EDPS.
- Gallup, G. (2012). Poner los sondeos al servicio de la democracia. En N. D'almeida, *La opinión pública* (págs. 51-67). Buenos Aires: La Crujía.
- Garrigou, A. (2007). *La embriaguez de las encuestas*. Santiago de Chile: LOM Ediciones.
- Germani, G. (1995). Surgimiento y crisis de la noción de opinión pública. En K. Young, *La opinión pública y la propaganda* (págs. 100-109). Buenos Aires: Paidós Studio.
- Goffman, E. (2012). *La presentación de la persona en la vida cotidiana*. Buenos Aires: Amorrortu.
- Gracia Pecourt, J. (2015). La esfera pública digital y el activismo político. *Política y Sociedad*, 75-98.
- Guttman, L. (1949). *The basis for scalogram analysis*. Chicago: Bobbs-Merrill, College Division.
- Habermas, J. (2009). *Historia y crítica de la opinión pública*. Barcelona: Editorial Gustavo Gili, S. A.
- Hall, B. (2012). Google critique: An application of depoliticization theory. *Technology in Society*, 251-255.
- Hand, D., Mannila, H., & Smyth, P. (2001). *Principles of Data Mining*. Cambridge: The MIT Press.
- Harto de Vera, F. (2006). Tipologías y modelos de democracia electrónica. *Revista de los Estudios de Derecho y Ciencia Política de la UOC*, 32-44.

- Hoang, T. y otros. (2013). Politics, sharing and emotion in microblogs. *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* , 282-289.
- Janowitz, M. (1975). Sociological Theory and Social Control . *American Journal of Sociology* , 82-108.
- Jansen, B. J., Zhang, M. S., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. *Journal of the American society for information science and technology*, 60 , 2169-2188.
- Keim, D. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics* , 1-8.
- Kosala, R., & Blockeel, H. (2000). Web Mining Resarch: A survey. *SIGKDD Explorations* , 1-15.
- Laclau, E. (2005). *La razón populista*. Buenos Aires: Fondo de Cultura Económica.
- Larsson, A. O. (2016). Online, all the time? A quantitative assessment of the permanent campaign on Facebook. *new media & society 2016, Vol. 18* , 274–292.
- Lazar, D. (2009). Life in the network: the coming age of computational social science. *Science* , 721–723.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of psychology* .
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies* , 1-167.
- Luhmann, N. (2009). *La política como sistema*. México: Universidad Iberoamericana.
- Marradi, A., Archenti, N., & Piovani, I. (2010). *Metodología de las Ciencias Sociales*. Buenos Aires: Cengage Learning Argentina.
- McCombs, M. (2006). *Estableciendo la agenda. El impacto de los medios en la opinión pública*. Barcelona: Paidós Comunicación.
- McCombs, M., & Shaw, D. (1972). The Agenda-Setting Function of Mass Media. *The Public Opinion Quarterly* , 176-187.
- Morozov, E. (30 de 01 de 2006). *¿Un robot para cada persona mayor?* Recuperado el 19 de 06 de 2016, de elpaís.com: [http://elpais.com/elpais/2015/11/05/opinion/1446722101\\_439342.html](http://elpais.com/elpais/2015/11/05/opinion/1446722101_439342.html)
- Nocera, P. (2008). Masa, público y comunicación. La recepción de Gabriel Tarde en la primera sociología de Robert Park. *Nómadas, julio-diciembre*, .
- Noëlle-Neuman, E. (1998). La espiral del silencio. Una teoría de la opinión. En J.-M. F. (comp), *El nuevo espacio público* (págs. 200-209). Barcelona: Gedisa Editorial.
- Noëlle-Neumann, E. (2003). *Las espiral del silencio. Opinión Pública: nuestra piel social*. Barcelona: Paidós.

- Olmo Oliver, P. (2005). El concepto de control social en la historia social: estructuración del orden y respuestas al desorden. *Historia Social*, nº 51 , 73-91.
- Papacharissi, Z. (2002). The virtual sphere: The internet as a public sphere. *New Media & Society February* , 9-27.
- Park, R. E. (1996). La masa y el público. Una investigación metodológica y sociológica. *Reis Nro 74* , 361-423.
- Park, R. E. (1925). *The City*. Chicago: Chicago University Press.
- Polat, R. K. (2005). The Internet and Political Participation: Exploring the Explanatory Links . *European Journal of Communication* , 435-459 .
- Raeder, T., & Chawla, N. (2010). Market basket analysis with networks . *Springer-Verlag* , 97–113.
- Reynié, D. (2012). Medir para reinar. En N. D'almeida, *La opinión pública* (págs. 39-49). Buenos Aires: La Crujía.
- Ross, E. A. (1901). *Social control, a survey of the foundations of order*. New York: Macmillan.
- Rubio Ferreres, J. M. (2009). Opinión pública y medios de comunicación. Teoría de la agenda setting. *Gaceta de Antropología* , 1-17.
- Sampedro Blanco, V., & Resina de la Fuente, J. (2010). Opinión pública y democracia deliberativa en la sociedad red. *Ayer Revista de Historia Contemporánea* , 139-162.
- Stieglitz, S. &.-X. (2012). Political communication and influence through microblogging--An empirical analysis of sentiment in Twitter messages and retweet behavior. *r. In System Science (HICSS), 2012 45th Hawaii International Conference* , 3500-3509.
- Tan, A.-H. (1999). Text mining: The state of the art and the challenges. *Proceedings of the PAKDD Workshop on Knowledge Discovery from Advanced Databases* . , 65-70.
- Tarde, G. (2011). *Creencias, deseos, sociedades*. Buenos Aires: Cactus.
- Thurstone, L. (1931). The measurement of social attitudes. *The Journal of Abnormal and Social Psychology* , 249-269.
- Tumasjan, A y otros. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM* , 178-185.
- Van Dijk, J. (2000). Models of Democracy and Concepts of Communication . En K. H. Dijk, *Digital Democracy* (págs. 30-53). Londres: Sage.
- Vosen, S., & Schmidt, T. (2013). Forecasting private consumption: survey-based indicator vs. Google trends. *Journal of Forecasting* , 30 (6), 565-578.

Warf, B. (2011). Geographies of global Internet censorship. *GeoJournal* , 1-23.

Westling, M. (2007). Expanding the public sphere: The impact of Facebook on political communication. *The New Vernacular* , 1-13.

Wilson, T. P. (1990). La sociología y el método matemático. En A. Giddens, *La teoría social, hoy* (págs. 489-514). Madrid: Alianza Editorial.

Wolton, D. (2008). Opinión pública y audiencia: en defensa de una reconstrucción teórica de los medios del público y de la comunicación. En N. D´almeida, *La opinión pública* (págs. 127-136). Buenos Aires: La Crujía.