

Un rato a solas con Google.

Por: Esteban Magnani. Revista Anfibia. 03/12/2017

Google o Pornhub nos conocen mejor que nosotros mismos. Son confesionarios a los que les contamos lo que a nadie más: fantasías sexuales, la mejor forma de dejar a tu pareja o el odio racial. Podemos olvidar qué hicimos, esconder nuestros secretos, pero en internet todo se almacena. ¿Hasta qué punto podemos tomar los datos de la vida online como indicadores confiables de la “realidad”? ¿Cómo se llevan las ciencias sociales y el Big Data?

En una [entrevista](#) de principios de 2017 el Dr. en Comunicación alemán Martin Hilbert decía que el Big Data convirtió a las ciencias sociales en la ciencia más rica en datos. “Antes tenías que negociar hasta con diplomáticos para que te prestaran una base de datos de cien filas por cien columnas. Y en las universidades hacían experimentos con quince alumnos de pregrado, que necesitaban créditos extra para pasar el ramo, todos blancos, todos de 18 años, y decían ‘miren, así funciona la psicología humana’”. Desde su punto de vista, gracias al enorme flujo de datos que proveen los dispositivos digitales, las ciencias “blandas” podrían finalmente alcanzar la densidad de las “duras” a fuerza de datos. Mientras la mayoría de los analistas [fracasaba](#) a la hora de explicar fenómenos como la reciente victoria de Trump (menos aún la [anticipaban](#)), el Brexit o el “no” al proceso de paz en Colombia, Hilbert señaló la capacidad de las herramientas informáticas para dirigir los mensajes con una eficiencia inesperada por los expertos, que con menos acceso a datos quedaron casi tan sometidos a sus [“burbujas”](#) como cualquier hijo de vecino.

Hilbert ve riesgo pero también un potencial enorme en esta posibilidad. “Nosotros nunca tuvimos datos, y por eso nunca funcionaban las políticas públicas. Y de la noche a la mañana, el 95% de los sujetos que estudiamos pasó a tener un sensor de sí mismo 24 horas al día. Los biólogos siempre dijeron ‘eso no es ciencia, no tienen datos’. Pero ellos no saben dónde están las ballenas en el mar. Hoy nosotros sí sabemos dónde están las personas, pero también sabemos qué compran, qué comen, cuándo duermen, cuáles son sus amigos, sus ideas políticas, su vida social”.

Gracias a los datos las ciencias sociales podría superar críticas como las del Ministro de Ciencia de Ciencia y Tecnología, Lino Barañao, cuando [reclamaba](#) ya en 2008: "...a mí me gustaría ver un cierto cambio metodológico; estoy tan acostumbrado a la verificación empírica de lo que digo, que a veces los trabajos en ciencias sociales me parecen teología". ¿Pueden (¿deben?) las ciencias blandas endurecerse para ascender en el escalafón científico?

Ponele Big Data a todo

Repasemos primero algo del poder de los datos que, cuando reúnen cantidad, velocidad en su flujo y herramientas de procesamiento, suelen llamarse Big Data. Hay quienes lo ven como un escalpelo capaz de llegar a donde nadie más lo hace. Es el caso del libro "[Everybody lies](#)" (Todos mienten) del economista, periodista del New York Times y ex-empleado de Google, Seth Stephen-Davidowitz (SSD de aquí en más) quien apoyado por grandes bases de datos (entre otros, las búsquedas de Pornhub, sitio porno que [publica sorprendentes estadísticas](#)) encontró respuestas que parecían imposibles.



SSD parte de algunos supuestos; uno de ellos es que buscadores como Google o sitios como Pornhub nos conocen mejor que nosotros mismos. Desde su punto de vista, el buscador funciona como un confesionario al que se le exhibe lo que a nadie más: fantasías sexuales, la mejor forma de abandonar a un novio o el odio racial. Uno puede olvidar qué hizo, mirar hacia otro lado con cara de “yo no fui”, pero en internet todo se almacena y lo devuelve a quienes, como SSD, saben aprovecharlo.

Por ejemplo, ¿son cuantificables los lapsus freudianos? ¿Es más frecuente tipear un error con connotaciones sexuales que uno sin sentido? SSD analizó 40.000 errores de tipeo de una base de datos (incluso los miles que fueron rápidamente corregidos) para encontrar las posibilidades estadísticas de que una letra reemplace a otra: por ejemplo, es común usar la “a” en lugar de la “s” porque están una al lado de la otra. Pero, ¿ese promedio estadístico se distorsiona para permitirnos sacar afuera lo reprimido? La probabilidad de un error de tipeo con connotación sexual es estadísticamente tan probable como cualquier otro.

SSD investiga otros temas sensibles como la homosexualidad para concluir, por medio de inteligentes cruces de datos, que está mucho más reprimida de lo que parecería a primera vista, al menos en los EE.UU. Según el autor, el buscador de Google funciona como una suerte de confesionario al que podemos decirle lo que a nadie más. También encontró en municipios donde Obama sacó porcentajes significativamente inferiores de votos en comparación con lugares sociodemográficamente similares, una correlación con un racismo oculto pero mensurable gracias a las búsquedas en Google.

“Todos mienten” busca inspirar a los escépticos, pero plantea algunos problemas metodológicos difícilmente aceptables si se busca hacer ciencia rigurosa. En primer lugar está el riesgo de tomar la vida online como un indicador confiable de la “realidad”. Veamos un último ejemplo para ilustrar esto: como es muy difícil obtener estadísticas confiables sobre la economía de muchos países, un investigador encontró una correlación entre la intensidad de las luces nocturnas detectadas desde los satélites y el desempeño económico nacional. De esta manera se podría saber en tiempo real cómo van las cosas en cada país. Pero, ¿cuánto tiempo pasaría desde que ese indicador se tome en serio hasta que los gobiernos comiencen a poner reflectores apuntando al cielo? De hecho, este fenómeno de intervención sobre “los datos” ya es [común](#) en Twitter, Facebook y demás sobre los

que se opera -incluso profesionalmente– para imponer un *trending topic* o disputar la narrativa sobre un hecho.

Digresión: es cierto que existen casos como el de [#SantiagoMaldonado](#) en el que una mayoría activa y militante logró imponer un tema por sobre el trabajo de los trolls pero sería arriesgado tomar esto como un síntoma de cierta neutralidad de las redes: si se festeja el éxito de la presión popular para imponer un tema particular, ¿debería también aceptarse la derrota cuando ganan las maquinarias pagas de internet?

El riesgo de no hacerse cargo

Tal vez lo más preocupante no sea la cuestión puramente epistemológica del asunto, si no el uso más pragmático y la racionalidad subyacente en el Big Data. En los casos mencionados ([Trump](#), Brexit, Colombia) las campañas apelaron a la construcción de perfiles muy segmentados de los votantes para encontrar los mejores mensajes capaces de interpelarlos para que voten. Gracias a la información que compartimos en las redes alimentamos las bases de datos a escalas nunca vista hasta permitir un mapa social hiperdetallado.

Un buen ejemplo es lo que muestran los distintos modelos de cartas que envié Cambiemos antes de las elecciones de octubre a numerosos ciudadanos. Una de las misivas comenzaba “¿Cómo estás? Primero quiero agradecerte por participar en las elecciones de agosto. Tus preocupaciones, críticas constructivas y sugerencias son muy importantes porque nos recuerdan que siempre se puede mejorar”. Otra versión decía “Primero quiero agradecerte porque juntos en agosto le dimos un gran respaldo al cambio que están en marcha en la Ciudad y en el país”. ¿Cómo supieron (o creyeron saber) la orientación del voto (secreto)? En realidad, no es tan difícil: posts en las redes, propios o de amigos, pueden resultar un excelente indicador. ¿El destinatario no publicó nada “político”? Mejor: los no interesados son los están menos aferrados a un partido y se los puede conquistar.

Este tipo de recursos de “lectura” y análisis en la red es lo que en el mundo del marketing se llama “social listening” y comprende análisis semántico, de circulación de mensajes y cuáles son los que más respuestas de distinto tipo generan: datos cada vez más refinados. Al complementarse con herramientas de publicidad como las que ofrecen Google o Facebook, es fácil segmentar también a quienes llegarán.

Difícilmente un sociólogo por muchos encuestadores que disponga podría procesar los millones de tuits, posts o búsquedas que sirven para decidir qué enviar a cada uno.



Pero lo más interesante probablemente sea la posibilidad de, por prueba y error, detectar por medio de algoritmos aquellos que efectivamente generan la respuesta deseada en mayor porcentaje. En el mundo digital ya no es necesario lanzarse al agua con una campaña multimillonaria y rezar para que no se hunda: se hacen “pruebas” que, según el caso, pueden alcanzar a miles de personas, casi mini campañas en sí mismas. Con esa información es posible elegir la mejor forma de

segmentar el uso de tal o cual palabra y generar un laboratorio para la experimentación antes del gran salto. En el lenguaje del marketing online este método se llama “A/B testing” y puede servir para elegir un discurso o indicarle al diseñador cuál de los dieciséis tonos de gris propuestos es el más “efectivo”. ¿Qué pasa cuando lo que genera atención está vinculado al racismo, la homofobia o un humor morboso? ¿Es válido simplemente por ser exitoso mirado desde la cerradura de la repercusión? Eso no lo responde el Big Data sino el que el da de comer.

Los cobayos de internet

En el mundo digital se producen muchos más datos de los que pueden llegar a procesarse humanamente. Sistemas como [Google Trends](#) peinan la red constantemente para indicarnos en qué está pensando “la gente”. Esta herramienta es utilizada, por ejemplo, para que un jefe de redacción encargue con urgencia a un periodista un artículo sobre “[Deyna Castellanos](#)”, el tema más buscado al momento de escribir esta nota. No importa que el artículo sea bueno, sino que se publique rápido y con un titular prometedor para generar [clicks](#) antes de que el tema se agote. Esta es una de las múltiples críticas que se le puede hacer al uso actual del Big Data: nos da lo que ya queríamos (faltaría averiguar por qué lo queremos) y no nos invita a cuestionarnos. Las decisiones editoriales basadas en Big Data tienden, por razones económicas, a reforzar lo que ya estaba y la sociedad no se ve desafiada a pensar más allá. ¿Es neutral que la gente lea solo lo que conoce? Si así fuera: ¿Cómo podría entonces pensar de otra manera o en algo nuevo? ¿Y si la lógica del Big Data estableciera que la mejor forma de mantener interesados a los estudiantes es jugar Clash Royal en las clases habría que aceptarlo sin más? Una limitación del Big Data es que dice lo que hay, pero no lo que podría haber.

Ricardo Diviani, Dr. en Ciencias de la Comunicación y docente en Epistemología de la Comunicación (UNR) entre otras materias y universidades, explica: “Yo creo que se puede ‘endurecer’ a las ciencias sociales con Big Data. La pregunta sería si es lo deseable o aconsejable para unas ciencias sociales que tengan vocación crítica y no sirvan, para decirlo de modo esquemático, a los poderes dominantes. Obviamente, una perspectiva crítica no significa “renegar” de este tipo de tecnologías sino disputar su sentido o, como se suele decir, resignificarlas. En este sentido, los datos estadísticos bien contruidos pueden ser utilizados para producir conocimiento significativos, como bien los señalaba Bourdieu, siempre y cuando no sean un fetiche que explica todo, o peor aún, intenten hacerse pasar como algo neutro”. Un

Big Data tal como está planteado resulta una herramienta del poder: “Ahora hay una fascinación por ‘cuantificar’ todo y quedarse en la respuesta del ‘cómo’ de los fenómenos sin pensar la importancia de los por qué. De ahí que noto cierto tufillo neopositivista en muchas de las ‘maravillas’ del Big Data. Considero que el fenómeno no puede ser analizado sin pensar los entramados de poder, la función ligada al mundo del capitalismo neoliberal y es desde estas consideraciones que sería interesante una apropiación del Big Data por parte de las ciencias sociales”.

El Dr. en Ciencias Sociales y docente (UBA) José Seoane coincide: “Las ciencias sociales o el pensamiento crítico se formulan desde otra perspectiva. Por supuesto que los datos, las evidencias empíricas son importantes y que las TIC plantean desafíos, pero la investigación social comienza con una pregunta que se construye con o desde la teoría y a partir de ahí construye sus datos. No es o no necesariamente es una técnica”. Plantear los datos como algo transparente es una postura fuertemente ideológica que esconde el proceso de selección, los recortes que se hacen y que implican necesariamente algún tipo de hipótesis.



Por otro lado, suelen “aplanar” la subjetividad a unas pocas dimensiones mensurables sobre las que se trabaja con insistencia para operar sobre la sociedad. El conductismo puede ser muy criticado como teoría psicológica, pero alcanza para explicar el poder de las estadísticas aplicadas a la conducta humana en los grandes números. Como señala Seoane, “La subjetividad es compleja y constituye todos esos momentos y sus contradicciones”. ¿Que alguien vea un juego sádico sexual por internet lo hace un sádico en la vida? Los sujetos, las sociedades, no es tan simple ni lineal.

Tener muchos datos no alcanza para controlar a la población; sería engañoso y

limitado considerar al Big Data como explicación suficiente para todo. Es cierto que puede ser un buen indicador de fenómenos no siempre visibles, pero no brinda un poder ilimitado para modificarlos a gusto o intervenir sobre ellos. Sería equivocado atribuirle toda la responsabilidad en la victoria de Trump, el Brexit o el referendun en Colombia: de hecho, los perdedores también lo usan.

El conocimiento detallado de la población puede servir para saber dónde y cómo encarar una campaña de vacunación de la forma más efectiva o proponer mejoras en la educación. El tema, como suele ocurrir, es quién define los objetivos y las interpretaciones de esos datos, es decir, la trama de poder subyacente como diría un cientista social. ¿Qué ocurre cuando alguien argumenta que se debe dejar hablar al Big Data? Cualquier dato requiere presupuestos mínimos que ayuden a ordenarlos y darles sentido; es en esos pliegues de supuesta neutralidad de “los datos” donde se esconden los mecanismos de poder. No existe algo así como una teoría objetiva y quien quiera simular que basa sus decisiones en el Big Data en realidad las estará delegando en intenciones ocultas o, como mínimo, inconscientes.

El Big Data es un fenómeno novedoso y, como tal, deslumbra. Sin embargo, la vieja y conocida teoría social, con todas sus dificultades y limitaciones para explicar un objeto de estudio que por su naturaleza es complejo, contradictorio e inabarcable, es capaz de explicar que usado por quienes lo proponen como la solución a todo más bien están vendiéndonos un caballo de Troya. Una vez más la pregunta no es tanto el qué sino, más bien, el cómo y el porqué.

[LEER EL ARTÍCULO ORIGINAL PULSANDO AQUÍ.](#)

Fotografía: Julieta De Marziani

Fecha de creación

2017/12/03