

“Previsión inmediata” de daños por desastres



De la “paradoja de la amistad” a la valoración del impacto
de un huracán utilizando datos de Twitter

MANUEL CEBRIÁN
Sociólogo informático

La historia en la que se basa este proyecto comenzó en 2010 cuando me encontraba en la UC de San Diego. Eran los primeros tiempos de la investigación utilizando conjuntos de datos de redes sociales. Para un joven becario, como era mi caso, la UCSD constituía un sitio fascinante para comenzar a realizar esta clase de trabajos. James Fowler, una de las mentes más brillantes en mi campo, la sociología informática, trabajaba en el campus. James y Nicholas Christakis—conocidos como

el “dúo dinámico” que descubrió el vínculo existente entre la obesidad y las relaciones sociales—publicaron uno de mis documentos favoritos en este campo. En dicho documento, demostraron que la paradoja de la amistad —el fenómeno por el cual tus amigos tienen más amigos que tú, descubierta por primera vez por Scott Feld en 1991 —podía utilizarse para detectar brotes de sarampión antes de lo que se lograba observando simplemente a la población general.

En el mejor de los casos, los actuales métodos utilizados para la detección de brotes contagiosos proporcionan información contemporánea acerca del desarrollo de una epidemia. Se sabe que, durante un brote, las personas situadas cerca del centro de una red social tienen más probabilidades de ser infectadas antes, en promedio, que aquellos que están situados en la periferia. Desafortunadamente, mapear individuos situados en una posición central a los que se podría monitorear para observar la evolución de la infección, suele resultar muy difícil. Proponemos una estrategia alternativa que no requiere la comprobación de una estructura de red global, es decir, basta simplemente con monitorizar a los amigos de individuos seleccionados al azar. Dichos individuos son conocidos por ocupar una posición más central. Para evaluar si un grupo de amigos de estas características pudiera facilitar una detección temprana, estudiamos un brote de gripe que se produjo en Harvard College a finales de 2009. Y realizamos un seguimiento a 744 estudiantes que eran miembros de un grupo de individuos escogidos al azar o bien de un grupo de sus respectivos amigos. A partir de diagnósticos clínicos, la progresión de la epidemia en el grupo de amigos se produjo 13,9 días antes (95 % C.I. 9,9–16,6) de la progresión en el grupo escogido al azar (es decir, la población considerada como un todo). Asimismo, el grupo de amigos mostró un tiempo de diagnóstico significativo ($p < 0.05$) en el día 16 de la epidemia, 46 días antes del pico de incidencia diaria en la población contemplada como un todo. Este método utilizado como sensor podría aportar un tiempo adicional significativo para reaccionar ante las epidemias en pequeñas o grandes poblaciones sometidas a vigilancia.

Christakis NA, Fowler JH (2010)

Sensores en redes sociales para la detección precoz de brotes contagiosos.

PLoS ONE 5(9): e12948. doi:10.1371/journal.pone.0012948

Durante un tiempo, estuve bastante interesado en probar el comportamiento de esta “paradoja de la amistad” en las nuevas estructuras sociales que ofrecían las redes sociales digitales. Esteban Moro estaba visitando la UC de San Diego durante el verano (procedente de la UC3M, España) y, junto con otro colega español que se hallaba en el campus, Manuel García-Herranz (procedente de la UAM, España), empezamos a “desguazar” algunos datos de Twitter procediendo a comprobarlos. Los resultados preliminares eran prometedores: así era, la paradoja de la amistad también tenía su reflejo en Twitter. Pero nos preguntamos qué significaba eso... ¿Acaso se trataba

de otra versión del fenómeno de la paradoja de la amistad en Internet? Decidimos comentarlo con James y Nicholas. Ellos sugirieron que, de la misma forma en que la gripe se extendió de una persona a otra en el campus de Harvard, la información también se extendía vía online y, quizás, la paradoja de la amistad también podría utilizarse para detectar un contagio global en Internet. Nos dispusimos a comprobar esta cuestión y, tras muchas horas de trabajo, pudimos demostrar que así era, también en Internet había la posibilidad de utilizar a los amigos como sensores para detectar brotes contagiosos globales/a escala.

Recientemente, se han llevado a cabo investigaciones centradas en la monitorización de datos online a escala global para mejorar la detección de epidemias, pautas de humor, movimientos en el mercado bursátil, revoluciones políticas, ingresos por taquilla, comportamientos de los consumidores y muchos otros fenómenos de importancia. No obstante, las consideraciones acerca de la privacidad y la gran amplitud de los datos disponibles online hacen que, rápidamente, la monitorización global resulte inviable y los métodos existentes no aprovechan al máximo la estructura de red local para identificar los nodos críticos para efectuar la monitorización. En este caso, desarrollamos un modelo de la diseminación epidémica de la información a una escala global, sobre una red social articulada públicamente y demostramos que un método simple puede producir no solo una detección precoz, sino que también puede advertir acerca de la existencia de brotes infecciosos. En este enfoque, escogimos al azar una pequeña fracción de nodos en la red y, luego, escogimos aleatoriamente a un amigo de cada nodo para incluirlo en un grupo para la monitorización local. Utilizando datos de 6 meses procedentes fundamentalmente de la esfera de Twitter, demostramos que este grupo de amigos tiene una posición más central en la red y nos ayuda a detectar brotes virales en relación con el uso de hashtags nuevos aproximadamente 7 días antes de lo que podíamos hacerlo utilizando un grupo escogido al azar de igual tamaño. Por otro lado, el método realmente funciona mejor de lo esperado debido a la estructura de red, ya que los actores que ocupan una posición marcadamente central son más activos y muestran mayor diversidad en la información que transmiten a otros. Estos resultados sugieren que la monitorización local no solo resulta más eficiente, sino también más efectiva y muestran que puede aplicarse al control de procesos contagiosos en redes a escala global.

García-Herranz M, Moro E, Cebrian M, Christakis NA, Fowler JH (2014)

Utilizar a los amigos como sensores para detectar brotes contagiosos a escala global.

PLoS ONE 9(4): e92413. doi:10.1371/journal.pone.0092413

Este descubrimiento produjo un significativo número de investigaciones académicas sobre aplicaciones online de la paradoja de la amistad (es el caso de Kristina Lerman et al. en la Universidad del Sur de California, Kay Axhausen et al. en ETH Zurich o Y. H. Eom et al. en la Universidad de Toulouse). Sus estudios, como ocurría con el nuestro, mostraban que eran capaces de detectar la diseminación viral en Internet, la “diseminación de la información”, pero ¿qué ocurre en el mundo real? ¿Podemos utilizar la paradoja de la amistad para detectar algo que está ocurriendo en el mundo físico?

Esta investigación se retomó en 2012, cuando estaba ocupando mi nuevo puesto en NICTA (ahora Data61, tras la fusión de NICTA y CSIRO), el centro más importante de Australia en investigación sobre ciencia de los datos. Allí colaboré con Pascal Van Hentenryck (ahora en la Universidad de Michigan), experto en gestión y mitigación de desastres en tiempo real. Pascal me reclutó para que trabajáramos juntos y exploráramos

si la información en tiempo real producida por las redes sociales resultaba suficientemente informativa como para mejorar los planes informáticos de evacuación en los que habíamos estado trabajando durante años. Mientras estábamos analizando esto, el huracán Sandy impactó en los Estados Unidos y viendo la tremenda devastación que produjo, nos preguntamos: ¿sería posible utilizar la paradoja de la amistad para detectar precozmente la destrucción que se está produciendo en la costa este de los Estados Unidos? Para esta misión, contamos con dos asociados postdoctorales con mucho talento: Caron Chen, doctorada recientemente en GIS, que se dedicó a obtener la mayor cantidad de datos posible acerca del huracán Sandy en Twitter y Yury Kryvasheyev, doctorado recientemente en física (modelos computacionales), que se centró en comprobar si las personas más conectadas enviarían un tuit antes que los demás cuando el huracán los alcanzara. Y, una vez más, funcionó, los sensores de redes sociales fueron capaces de enviar advertencias tempranas durante el impacto del huracán Sandy:

El flujo de la información durante eventos catastróficos constituye un aspecto crítico de la gestión de desastres. Las plataformas modernas de comunicación, particularmente las redes sociales online, proporcionan una oportunidad de estudiar dicho flujo y extraer sensores de advertencia precoz, mejorando así el grado de preparación y respuesta ante las emergencias. El rendimiento del método de sensores de las redes sociales, basado en propiedades topológicas y propiedades del comportamiento que surgen a partir de la “paradoja de la amistad”, se estudia en este caso atendiendo a más 50 millones de mensajes en Twitter publicados antes, durante y después del huracán Sandy. Comprobamos que las diferencias en la centralidad de la red de los usuarios realmente se traduce en una ventaja de conciencia moderada (hasta 26 horas) y que la geolocalización de los usuarios dentro o fuera del área afectada por el huracán juega un papel significativo a la hora de determinar la escala de dicha ventaja. La respuesta emocional parece ser universal, independientemente de la posición que ocupa en la topología de la red y muestra patrones característicos y fácilmente detectables, abriendo la posibilidad de implementar una técnica simple de “percepción del sentimiento” que puede detectar y localizar desastres.

Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P, Cebrian M (2015)

Rendimiento de los sensores de red social durante el huracán Sandy

PLoS ONE 10(2): e0117288. doi:10.1371/journal.pone.0117288

En aquellos momentos (2014 -2015) Twitter ya se identificaba como una plataforma que resulta bastante útil cuando se producen desastres naturales o de carácter tecnológico. Los investigadores estudiaron a fondo el papel de los servicios online a la hora de facilitar el flujo de información relacionado con las emergencias, detectando eventos online y recaudando fondos de manera colaborativa para mitigar los daños. Una vez más, con raras excepciones (Guan y Chen, 2014, Uso de datos de redes sociales para comprender y evaluar los desastres), la mayor parte de los estudios hasta el momento estaban relacionados con la información y conformaron el campo de lo que hoy se denomina informática de crisis. Ahora, el siguiente paso no solo consistía en detectar información acerca del desastre, sino determinar si podíamos inferir la magnitud del daño que se estaba produciendo en las distintas comunidades utilizando exclusivamente datos procedentes de Twitter. Nuestro siguiente movimiento fue una apuesta arriesgada, ya que pensábamos que esto iba mucho más allá de lo que podíamos extraer de Twitter. Una cosa es la información, respecto a la cual sabemos que las redes sociales se comportan bien, pero ¿qué ocurre con los daños producidos en el mundo real? Se trata de un nivel de impacto mucho más elevado y no estábamos seguros de que todos los vínculos estadísticos pudieran sobrevivir al ruido presente en Twitter.

Por tanto, establecimos en el transcurso del análisis la distribución espacial de los tuits durante la supertormenta Sandy. Los resultados preliminares de Yuri indicaron que la actividad online normalizada (mensajes políticos sobre el tema per cápita) tienen una relación directa con la proximidad a la zona del desastre, con un marcado incremento en las cercanías del punto de impacto. Esto nos dio la esperanza de que la fuerza impulsora que subyacía al tuit fuera una combinación de factores que hacen que el hecho de tuitear sea relevante: ¿hasta qué punto resulta peligroso al evento, cuál es su nivel de cercanía, hasta qué punto sus manifestaciones resultan obvias (vientos, marejada ciclónica, etc.) y cuál es el nivel de severidad de las consecuencias? Consideramos que la relación entre la



actividad virtual y los daños producidos en el mundo real podía detectarse.

Las mediciones del daño en el mundo real son difíciles de obtener. Afortunadamente para nosotros, Yury Kryvasheyev entró en contacto con una especialista en daños ocasionados por desastres. Se trata de Colleen O'Dea (de [NJ Spotlight](#)), que prestó una gran ayuda a la hora de proporcionar consejos sobre cómo obtener datos relacionados con los daños producidos por el huracán Sandy a partir de diversas agencias (FEMA, el Departamento de Banca y Seguros del Estado de Nueva Jersey y el Departamento de Servicios Financieros del Estado de Nueva York).

Posteriormente, decidimos incorporar a antiguos miembros del equipo (Esteban Moro, James Fowler) para ayudar a evaluar los análisis estadísticos que consideramos más severos que aquellos que figuraban en documentos anteriores. Para procesar los datos, James también captó la colaboración de un joven científico climático, Nick Obradovich, que estaba realizando su doctorado en la Institución Oceanográfica Scripps.

Teniendo todos los ingredientes a mano, procedimos a comparar dos distribuciones espaciales: la actividad online normalizada en comparación con los daños producidos por el huracán per cápita. Y, demostrando nuestra suposición intuitiva, la correlación resultó bastante significativa (alcanzado un valor de 0,6 en la correlación de Pearson). Los resultados fueron matizados, por ejemplo, en relación con los daños, la actividad posterior al desastre resulta más predictiva que el aluvión de tuits durante el momento álgido del desastre. Asimismo, una resolución espacial del análisis afecta a la fuerza de las correlaciones observadas. Finalmente, la metodología de la evaluación de daños (que se tomó en cuenta) así como las subvenciones individuales de FEMA están moderadamente correlacionadas con la actividad online, pero conjuntamente con los datos sobre seguros, la relación resulta mucho más fuerte (sugiriendo que cuanto más completa es la imagen del daño monetario, más estable resulta la relación). Fundamentalmente, verificamos nuestros descubrimientos en relación con



otros eventos acaecidos en los Estados Unidos y hallamos resultados similares en relación con escalas espaciales y tipos de amenazas: grandes tormentas y tornados, inundaciones, deslizamientos de tierras y terremotos.

En resumidas cuentas, dejando a un lado los detalles más elaborados, el resultado más importante es una excitante solución prácticamente a coste cero que permite realizar la “previsión inmediata” de los daños producidos por desastres. No solo eso: nuestro método también podría ayudar en la evaluación posterior de la catástrofe, ajustando los informes oficiales y las herramientas predictivas (basadas, por ejemplo, en la trayectoria del huracán y las simulaciones de fragilidad) con los datos de las redes sociales y valorando el impacto, la gestión y la efectividad de las medidas implementadas en el transcurso del desastre. Mediante la monitorización del huracán utilizando redes sociales, los gestores de emergencias pueden adoptar acciones en aquellas áreas en las que las redes sociales mostraron una actividad anormalmente alta. En algunos países, menos preparados para afrontar desastres, hay un considerable interés en explorar las redes sociales en lo que respecta a la respuesta ante emergencias. Nuestros estudios constituyen pasos positivos en esta dirección: Indican que las redes sociales capturan algunos elementos que son difíciles de obtener mediante medios tradicionales. ■