

[Big data, espionaje y movimientos sociales](#)

Enviado por marta el Mar, 10/29/2013 - 18:30

Sección principal:

[Saberes](#)

Cuerpo:

En el mundo tecnológico, la moda [Big data](#) ha sustituido a otras modas como el [cloud computing](#) o el [BYOD](#). Sin embargo, son muchos los malentendidos sobre lo que es realmente Big data y cual es su utilidad. Como leí hace tiempo, Big data es como el sexo adolescente: "Todo el mundo habla de ello, nadie sabe realmente lo que es, pero como todos creen que los demás ya lo están haciendo, tú también dices que lo estás haciendo". En el mundo del Big data para espionaje y control de los movimientos sociales, no es muy diferente.

¿Qué es Big data?

Big data es la capacidad de extraer información relevante de conjuntos de datos que cumplen las tres V: gran volumen de datos, alta velocidad de generación y en gran variedad de formato. Sin embargo, lejos de lo que nos muestra la ficción, **no nos permite detectar patrones que se salgan de la media, sino tendencias**; es decir, lo contrario. Big data no es capaz de encontrar una aguja en un pajar, sino que sólo nos valdría para decirnos qué longitud y tonalidad tienen la mayoría de los trozos de paja a base de analizarlos todos y detectar tendencias. Nos dice cual es el estado de la moda, no si una persona está fuera de ella.

Big data no es capaz de encontrar una aguja en un pajar, sino que sólo nos valdría para decirnos qué longitud y tonalidad tienen la mayoría de los trozos de paja a base de analizarlos todos y detectar tendencias.

Lo mejor es analizarlo con el ejemplo que lo puso de moda, las últimas elecciones presidenciales de EE UU y como Nate Silver fue capaz de predecir los resultados con un margen del 0,1% de error. ¿Cómo lo hizo? Utilizó las encuestas electorales que se iban publicando, las cuales tienen margen de error en torno al 2%, siempre y cuando la metodología sea correcta y no haya preguntas trampa. Lo que hizo Silver fue ponderar esas encuestas en función de las preguntas que hacían y la fuente y sumarlas, lo cual le permitió acertar los resultados de 49 de los 50 estados (excepto en Indiana, donde el 0,1% significó la derrota de Obama). Lejos de significar la muerte de las encuestas (ya que el método de Silver las necesita y cuantas más mejor), lo que mostró fue la capacidad de abordar su análisis de una forma totalmente diferente al habitual. Meses después se supo que **el equipo de Obama utiliza Big data para analizar las encuestas y las redes sociales para afinar el discurso presidencial**. Según su equipo electoral, llegaron a tomar 70 decisiones diarias basadas en Big data.

Nadie duda ya de la capacidad de Big data para detectar tendencias, e incluso han aparecido varios colectivos que están empezando a utilizar Big data como forma de activismo. Así por ejemplo, el colectivo [DatAnalysis15M](#) utiliza datos capturados de las redes sociales para "predecir" el resultado de las convocatorias así como para estudiar los sentimientos volcados los días previos para saber qué estados de ánimo y formas de comunicación parecen ser más efectivas para conseguir movilizaciones. Otro muchos para analizar la prensa o incluso el Boletín Oficial del Estado.

Un mecanismo complementario

Durante las últimas semanas, según se iban haciendo públicas las revelaciones del espionaje global de la NSA, el Gobierno ha defendido estas escuchas amparándose en el uso de Big data para detectar ataques terroristas como un nuevo 11S. Sin embargo, como hemos dicho, Big data nunca nos dará esta información. La mayoría de **acontecimientos como el 11S o el 11M son 'cisnes negros', es decir, patrones únicos imposibles de localizar**. Además, aún con un margen de error del 0,1% como el algoritmo de Nate Silver, hablaríamos de un falso positivo cada 1000 casos. Es decir, que aplicado a la población del Estado, detectaríamos 47.000 falsos terroristas. No parece

la mejor opción para pillar a los terroristas de verdad.

¿Por qué se espía ese volumen de comunicaciones? Para almacenarlas y buscar rastros de los sospechosos que se descubran de otras formas. Ilegal, pero ni mucho menos un Gran Hermano.

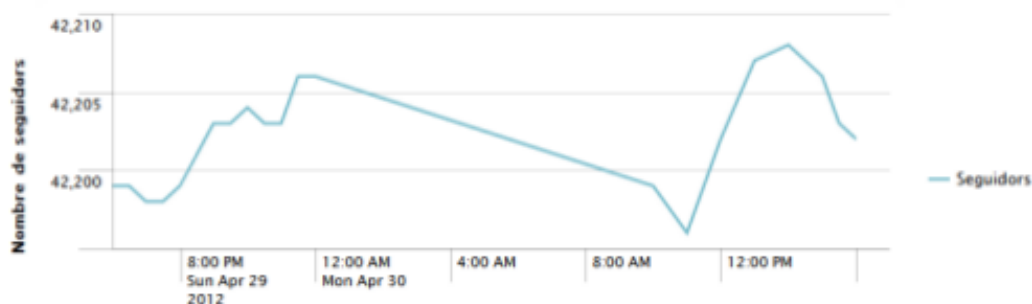
Si el Big data no permite hacer eso, ¿por qué entonces se espía ese volumen de comunicaciones? Por una sencilla razón: almacenarlo. Para poder buscar los rastros de aquellos "sospechosos" que descubran de la forma tradicional. Ilegal, por supuesto, pero ni mucho menos un Gran Hermano que busque patrones terroristas. Una vez identificado un nuevo objetivo, les permite saber qué estaba haciendo esa persona en un momento pasado concreto y saber con quién se relacionó. Algo que el sistema informático de la Hacienda española también usa de forma idea similar: analizando las cuentas y movimientos de las personas que han tenido relación con alguien que ya haya sido detectado como defraudador, aplicando la teoría de que el fraude funciona como una red. Cuando la NSA capturó 60 millones de datos de llamadas en España entre diciembre de 2012 y enero de 2013 **no buscaba capturar en sí el contenido de las llamadas, sino mapear las redes personales de determinadas personas para detectar otras personas de interés** a os que espiar con mayor atención.

¿Y el [análisis que realizó CESICAT para los Mossos y que fue publicado por Anonymous](#) hace unos días? La realidad es que en este caso no podemos hablar ni de Big data ni de 'small data', sino de no entender lo que realmente están haciendo. Me temo que es bastante sintomático **que los Mossos d' Esquadra no sean capaces de monitorizar un hashtag de Twitter en tiempo real, y tengan que pedirles a dos expertos en seguridad informática que les manden un resumen** cada seis horas. Un análisis que por otro lado no puede pecar de más simplista ya que las personas encargadas de elegir los tuits no tienen ningún conocimiento de los movimientos sociales. No me sorprendería que después de esa investigación se haya descartado seguir haciendo este tipo de monitorización por su incapacidad de extraer información, lo cual tampoco debería sorprender de lo que se ha terminado en llamar "la TIA catalana".

oscartaxibcn	1203	29/04/2012 23:00
axeбра	991	29/04/2012 23:00
aritzcirbian	926	30/04/2012 12:00
patxigu	722	30/04/2012 6:00
CatEmprenyat	585	30/04/2012 6:00
1demaig	466	29/04/2012 23:00
adestemps	411	29/04/2012 23:00
Freedom For Cat	250	30/04/2012 6:00
laboral15m	242	29/04/2012 23:00
maria_kny	104	30/04/2012 12:00
carucha_22	101	30/04/2012 6:00
jvrgarrido	57	30/04/2012 6:00

Evolució del Top 3 perfils amb més seguidors

Evolució seguidors del perfil @Acampadabcn durant l'operatiu



Pie de foto:

Fragmento del informe de la Fundació CESICAT (Centre de Seguretat de la Informació de Catalunya) en el que se puede observar que el autor de este artículo fue monitorizado desde el 29 de abril de 2012.

Temáticos:

[Big data](#)

[Control social](#)

Geográficos:

[Catalunya](#)

Artículos relacionados:

[El laberinto de la identificación en internet](#)

[Útiles legales para proteger tus datos](#)

[Usar \(o no\) capucha digital](#)

[Mortadelo y Filemón han empezado a seguirte en Twitter](#)

[El cíclope vs. la multitud distribuida](#)

Edición impresa:

Licencia:

[CC-by-SA](#)

Posición Media:

Cuerpo del artículo

Compartir:

Tipo Artículo:

Normal

Autoría:

[Txarlie Axebra](#)

