



> **ENTREVISTA**

## Alessandro Vespignani

Profesor de la Universidad Northeastern en Boston, donde dirige el Laboratorio de Modelado de sistemas biológicos y socio-técnicos. / Por **Elena Soto**

# «Los datos son la nueva minería de oro y de diamantes»

Una de sus áreas de investigación es el modelado de la distribución espacial de las epidemias. Vespignani ha sido uno de los ponentes del IWSOS 2013, un evento organizado por el Instituto de Física Interdisciplinar y Sistemas Complejos (IFISC, CSIC-UIB), dedicado a la auto-organización de redes y sistemas en red, incluyendo los sistemas tecno-sociales.

**Pregunta.**— 'Big data', minería de datos ¿es la información el nuevo filón de metales preciosos del siglo XXI?

**Respuesta.**— Es la minería de oro y de diamantes. Cambiará completamente el campo científico y sobre todo tendrá un fuerte impacto en el área de las ciencias sociales. La idea es que con la minería de datos podamos finalmente desarrollar la tecnología de modelos predictivos para la comprensión de los sistemas sociales, de la economía, de la difusión del conocimiento científico, de la propagación de enfermedades o de la evolución de patrones de comportamiento, entre otros. El objetivo final es realizar una meteorología de la sociedad que permita predecir lo que puede pasar. Es decir, desarrollar predicciones cuantitativas que posteriormente se pueden contrastar.

**P.**— ¿Qué tiene en común el cerebro, Internet, las redes eléctricas y las tróficas?

**R.**— El disponer de grandes cantidades de datos nos ha permitido descubrir propiedades complejas como las fluctuaciones a gran escala y heterogeneidades en muchas de las redes que hasta hace poco solo se consideraban como un mero conjunto de puntos y cone-

xiones casuales. Muchos sistemas, desde Internet al cerebro, pasando por las redes eléctricas, de tráfico o de los ecosistemas, se pueden representar como grandes redes complejas. Se han producido dos grandes revoluciones; una es el poder disponer de grandes cantidades de datos, 'big data': y la otra es la ciencia de sistemas complejos que ha matado el mito de la hormiga reina; es decir, que el funcionamiento del hormiguero no puede reducirse al comportamiento y las propiedades de cada uno de sus miembros, sino al comportamiento emergente derivado de la interacción de todos.

**P.**— En 2009, el virus de la gripe H1N1 se propagó desde México a diferentes partes del mundo, usted participó en la construcción del modelo computacional GLEAM ¿Cómo puede predecirse la expansión de una epidemia?

**R.**— Cada enfermedad tiene su propio modelo. Es una infraestructura computacional complicada que integra tres factores. El primero emplea una base de datos de población —los 7.000 millones de habitantes del planeta—, el segundo se centra en cómo se mueve la gente —bases de datos de avión, flujos de desplazamientos diarios estimando la movilidad humana dentro y entre los países—, y por último, se representa el comportamiento de la enfermedad, teniendo en cuenta los tiempos de incubación y de transmisión. Este aspecto tiene mucha importancia porque en principio, aunque los síntomas no se manifiesten, la persona puede haber viajado varios días con el virus. A partir de aquí podemos informar, por ejemplo, de cuándo se pro-



Alessandro Vespignani, de Northeastern University, en Boston.

ducirá el pico máximo de incidencia de la pandemia.

**P.**— ¿Acertaron en sus predicciones?

**R.**— Publicamos los resultados tres meses antes de finalizará la epidemia y nuestras previsiones coincidieron con las reales. Se pensaba que el pico estaría en los meses de enero y febrero, pero en nuestro modelo el número máximo de casos se producía en la primera semana de noviembre, que fue lo que sucedió. Las vacunas llegaron tarde, en el mes de diciembre, cuando el pico ya había comenzado a descender, lo que originó un elevado coste económico. Los modelos computacionales son instrumentos valiosos y muy útiles para los gestores, ya que les pueden ayudar a tomar decisiones proporcionándoles una información cuantitativa. Otro dato interesante que descubrimos fue que era ineficaz suspender los vuelos, ya que lo único que se conseguía era retrasar la epidemia una o dos semanas.

**P.**— ¿Qué propiedades complejas se han descubierto en muchas redes al analizar los big data?

**R.**— El pensamiento lineal no funciona en sistemas complejos. Si realizas una pequeña perturbación en un sistema lo lógico es pensar que la respuesta también será pequeña

y que si la perturbación es grande la respuesta será grande. Pero en los sistemas complejos no es así, una mínima perturbación puede crear una gran avalancha. Es algo que parece contrario a la intuición. Un ejemplo es el cierre de los aeropuertos en caso de la epidemia. Parece una medida efectiva y sin embargo no lo es.

**P.**— Se sabe que la minería de datos es de gran utilidad en el marketing ¿en qué otras áreas menos conocidas se emplea?

**R.**— En la política, la ciencia, la medicina, la planificación urbana, la organización social o la gestión del tráfico; los campos son innumerables. El análisis de datos de una plataforma como Twitter, por ejemplo, permite hacer sondeos electorales, conocer la nacionalidad de los turistas que visitan un país o saber qué lenguas se emplean en un determinado territorio, todo sin necesidad de encuestas. Es una tomografía de la sociedad por capas. En los próximos años podrán realizarse previsiones sociales muy acertadas. La pregunta es ¿qué puede suceder si se dispone de esta información y se hace pública? ¿Hasta qué punto puede haber retroalimentación en el sistema? Sin duda cambiará el modo de entender la sociedad y de tomar decisiones.



## RECERCA



### Ifisc Reuneix 35 investigadors de tot el món

► L'Institut de Física Interdisciplinària i Sistemes Complexos (IFISC, CSIC-UIB) ha organitzat la setena edició de l'International Workshop of Self-Organizing Systems (IWSOS 2013), en el qual participen 35 investigadors d'arreu del món per debatre des d'un enfocament multidisciplinari i innovador sobre l'autoorganització en xarxes i en els sistemes en xarxa, com són les xarxes de transports o de telefonia i les xarxes financeres, socials i biològiques.