

# BIG DATA

Javier Gutiérrez Puebla  
Departamento de Geografía Humana  
Universidad Complutense de Madrid



tGIS

transporte infraestructuras y territorio  
Grupo de Investigación de la Universidad Complutense de Madrid

# BIG DATA

- Datos generados por máquinas: Redes de sensores.
- Contenidos generados por los usuarios: La web 2.0: opiniones, valoraciones, comportamientos (trip advisor, trending topics...)
- Producción de datos masivos a un ritmo antes insospechado. El valor de los datos. El petróleo de la revolución de los datos.



# TECNOLOGÍA Y DISPOSITIVOS QUE PRODUCEN O ALMACENAN DATOS DE NUESTRA ACTIVIDAD COTIDIANA

1/6/14. Videovigilancia: las imágenes pueden ser interceptadas.

2. Contadores de luz y termostatos: dan información de hábitos.

3 /4. Televisores inteligentes y consolas: incorporan cámaras y micrófonos.

5. Controles biométricos de entrada y salida.

7. Monitorización remota en el trabajo: capturas de pantalla del trabajador para medir la productividad.

8. Bases de datos personales: pueden contener datos fiscales y de salud de los clientes.

9. Sensores de conteo de personas: monitorean flujo de compradores y los tiempos de compra.

10. Tarjetas de fidelización: a cambio de descuentos, crean perfiles del comprador.

11. Ibeacons: envían ofertas a móviles cercanos.

12. Wifi gratuito: se puede ofrecer a cambio de acceso al perfil de Facebook.

13. Abonos de transporte público: tarjetas recargables que producen datos de desplazamientos.

14. Redes de bicicletas públicas: registro de trayectos.

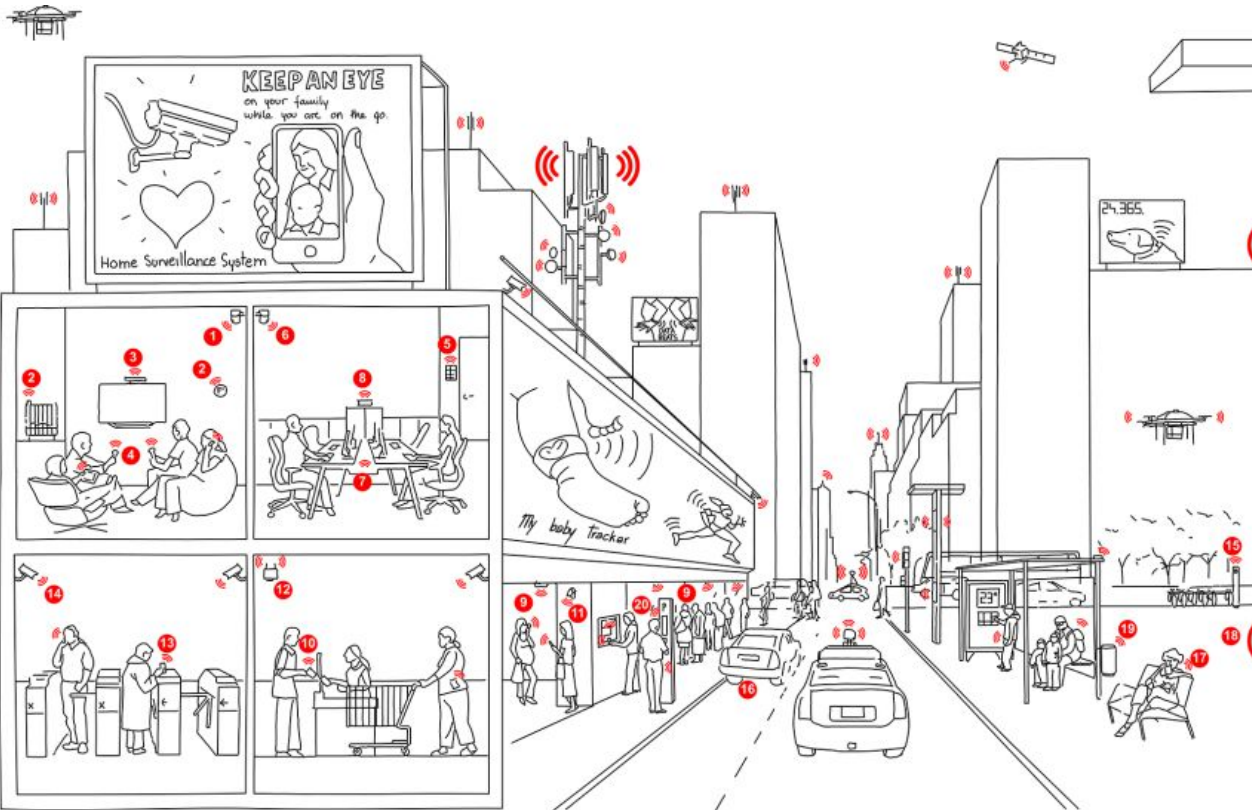
15. Coches: hay sistemas para registrar las matrículas.

16. Telefonía móvil: permite geolocalizar.

17. Cámaras térmicas y sensores sonoros: miden flujo de peatones y niveles de ruido.

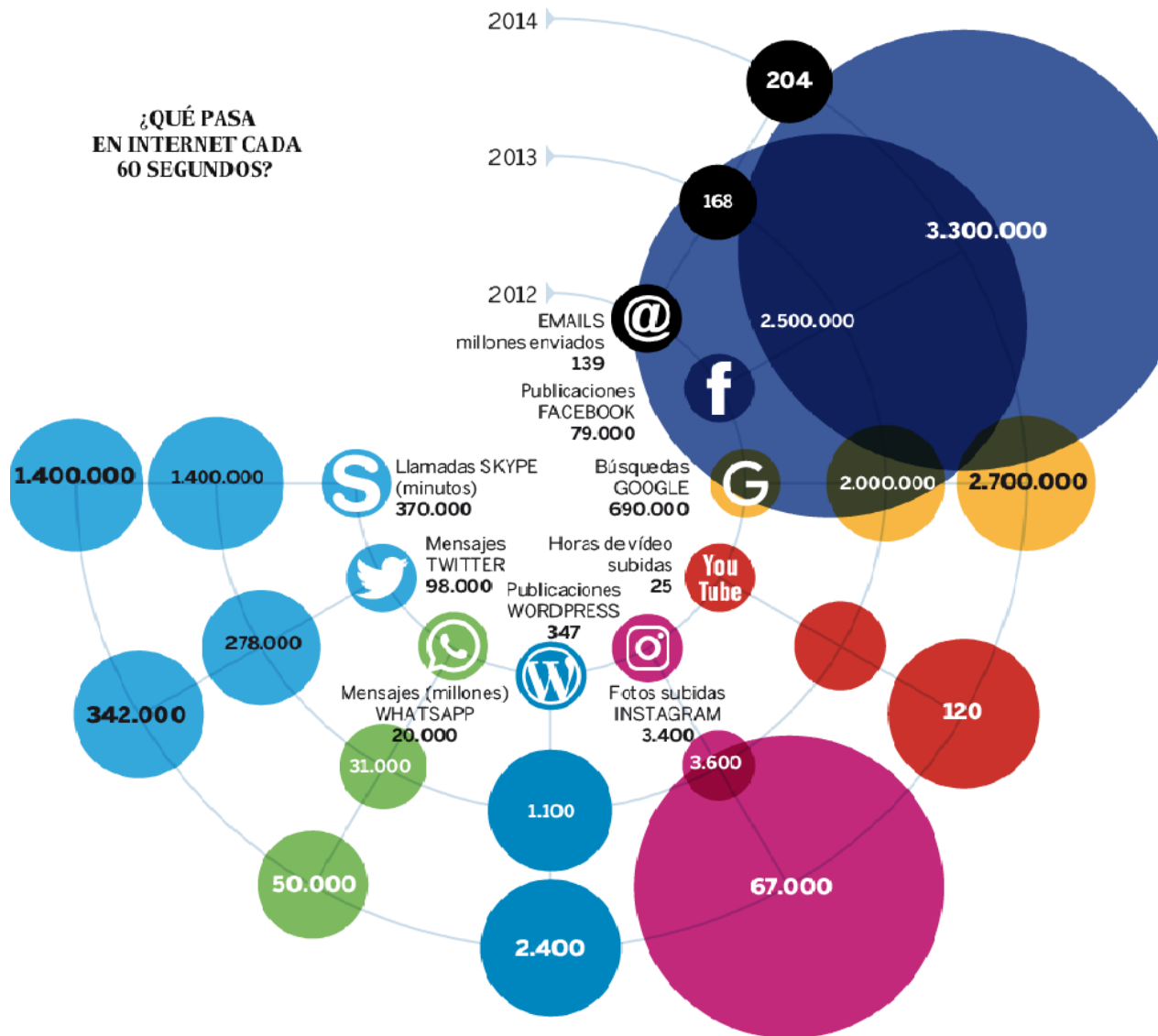
18. Mobiliario urbano que detecta presencia de peatones.

19. Sistemas de parking: el pago con tarjeta de plazas azules y verdes genera datos del usuario.



# INTERNET Y BIG DATA

¿QUÉ PASA  
EN INTERNET CADA  
60 SEGUNDOS?



# EL VALOR DE LOS DATOS: FACEBOOK

## ¿Cuánto gana Facebook por usuario?

Ingresos medios anuales por usuario según áreas geográficas\*



@Statista\_ES

Noviembre 2016

\* Calculado según los ingresos del tercer trimestre de 2016.

Fuentes: BuzzFeed, Facebook

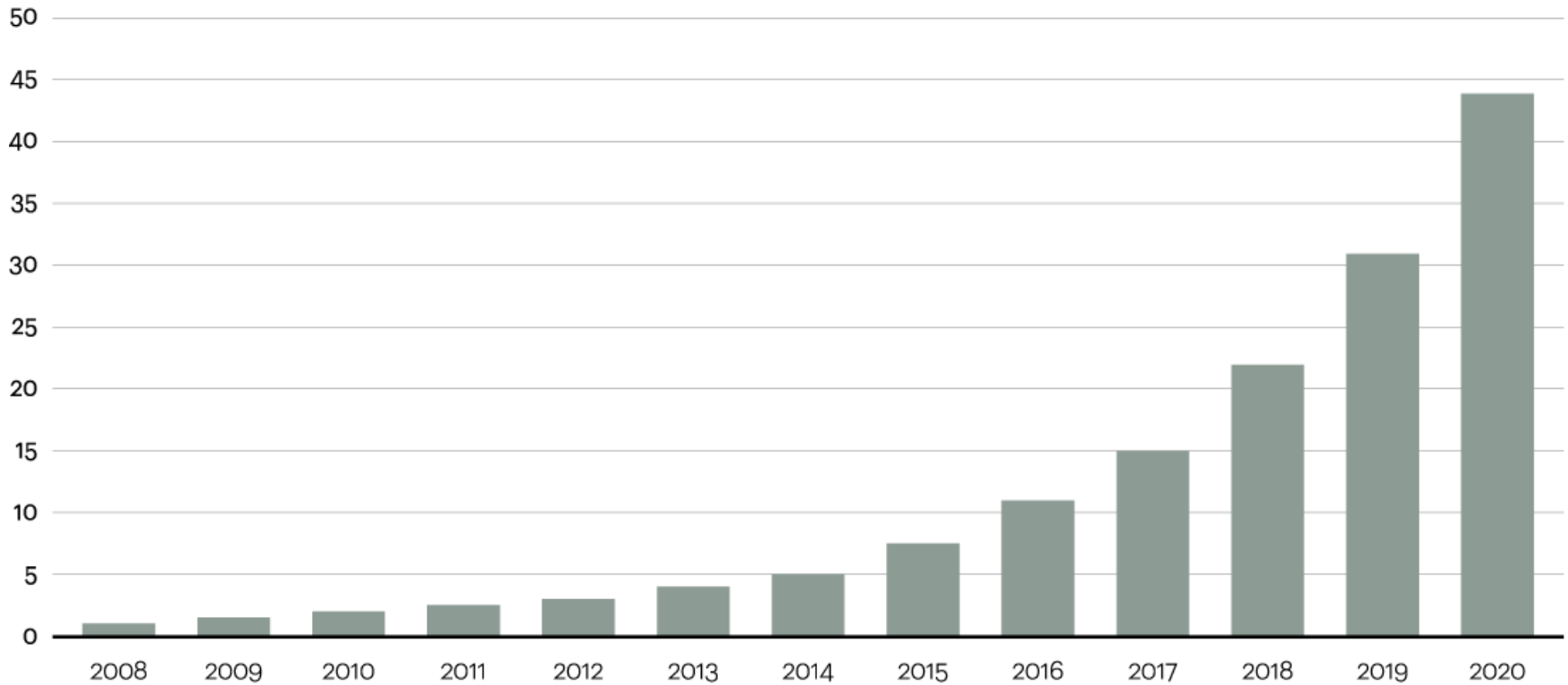
ABC **statista**

# LA EXPLOSIÓN DE LOS DATOS

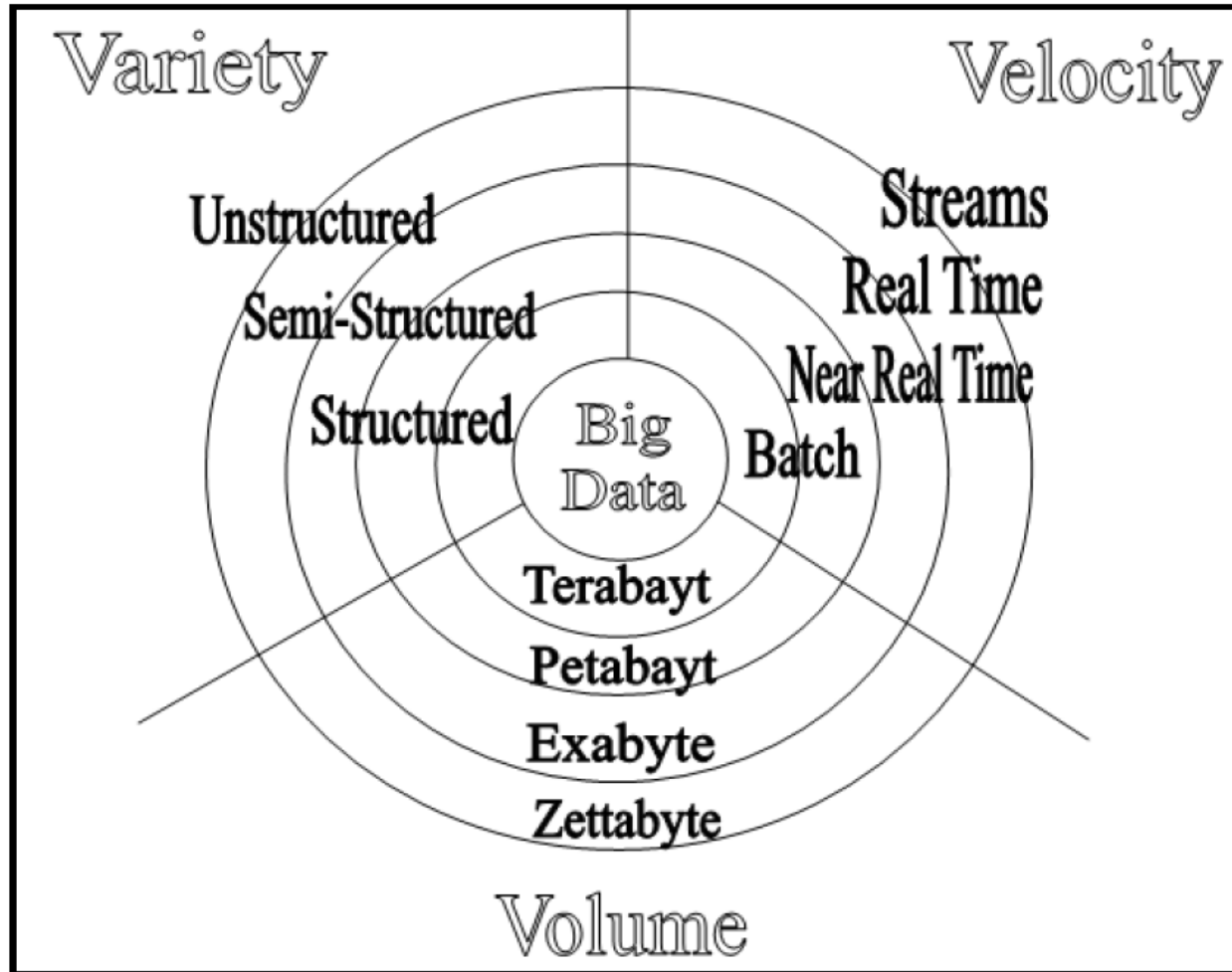
Figure 1

**Data is growing at a 40 percent compound annual rate, reaching nearly 45 ZB by 2020**

## Data in zettabytes (ZB)



## CARACTERÍSTICAS DE BIG DATA: LAS 3 V



Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In Collaboration Technologies and Systems (CTS), 2013 International Conference on (pp. 42-47). IEEE.

# ALMACENAMIENTO Y PROCESO DE BIG DATA ¿EN LA NUBE?: INFRAESTRUCTURAS Y REDES



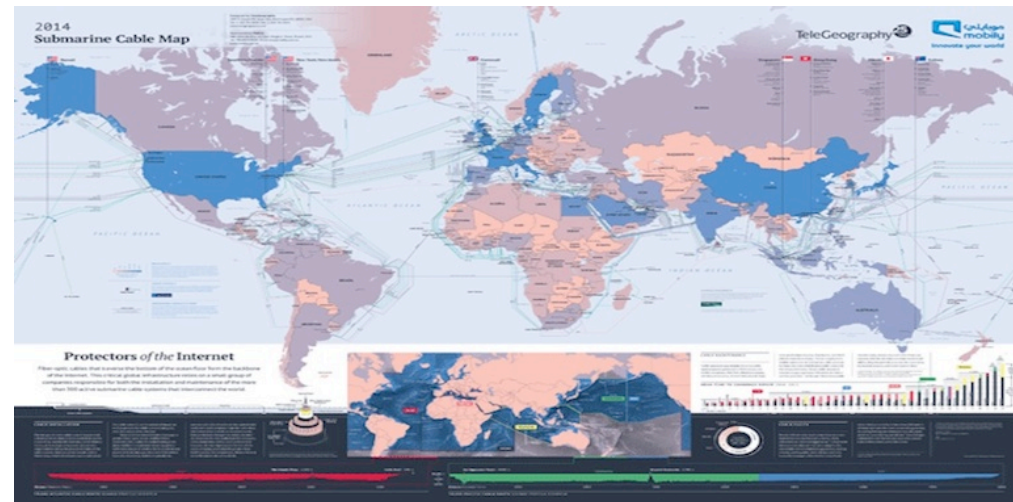
2% del consumo mundial de energía



Servidores de Facebook en Suecia



Vista aérea de los data centers de Quincy, WA



Cableado submarino y vulnerabilidad de la red



# EL PODER DE LOS DATOS EN LAS EMPRESAS

## El mundo comienza a girar alrededor de los datos



Data are becoming the new raw material of business.

— Craig Mundie —

**Craig Mundie**  
Head of Research and Strategy at Microsoft

**Tim Berners-Lee**  
Inventor of the World Wide Web



Data is a precious thing and will last longer than the systems themselves.

— Tim Berners-Lee —

AZ QUOTES

# EL PODER DE LOS DATOS EN LA INVESTIGACIÓN: EL CUARTO PARADIGMA

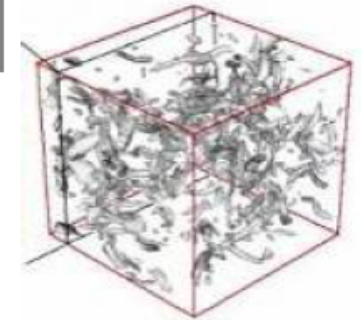
\* Thousand years ago:  
science was **empirical**  
describing natural phenomena



\* Last few hundred years:  
**theoretical** branch  
using models, generalizations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{4\pi G\rho}{3} - K\frac{c^2}{a^2}$$

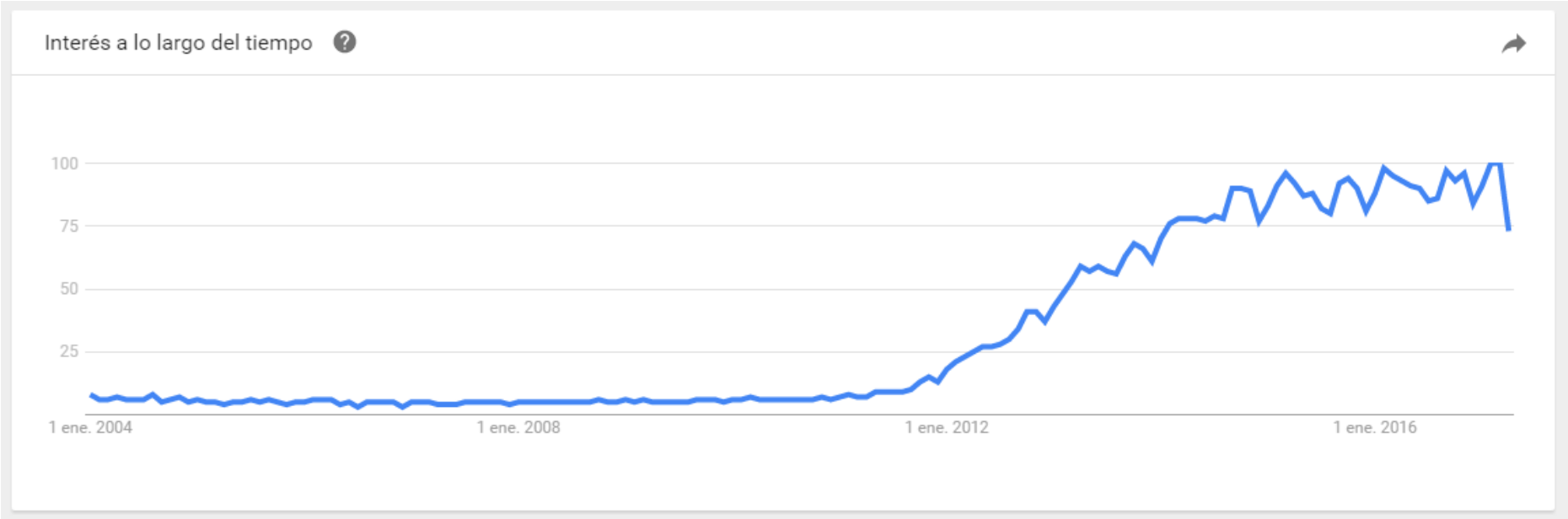
\* Last few decades:  
a **computational** branch  
simulating complex phenomena



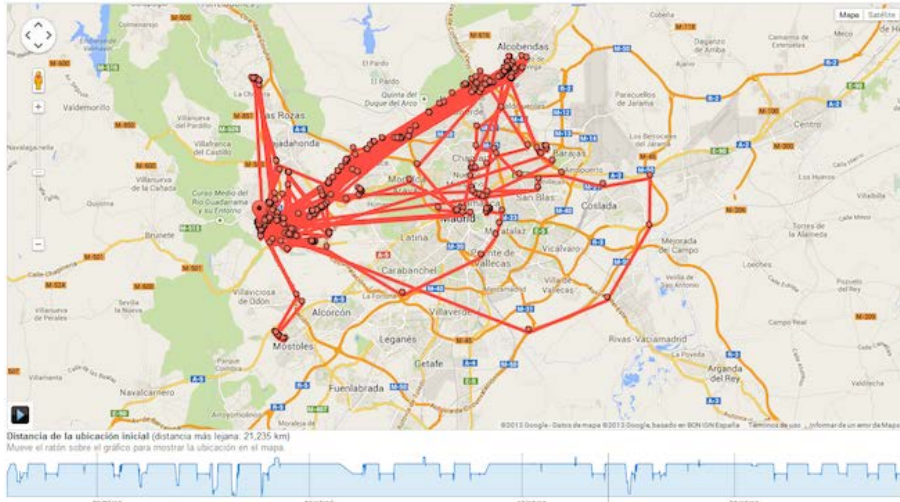
\* Today:  
**data exploration (eScience)**  
unify theory, experiment, and simulation  
Information/Knowledge stored in computer  
Scientist analyzes database / files  
using data management and statistics



# BÚSQUEDAS CON GOOGLE: GOOGLE TRENDES

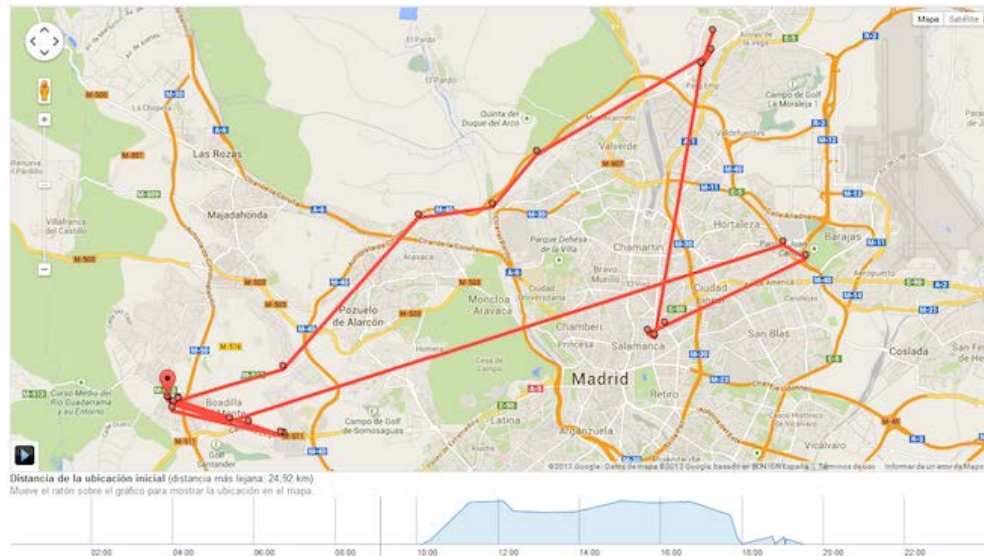


# Google Maps: LOCATION HISTORY

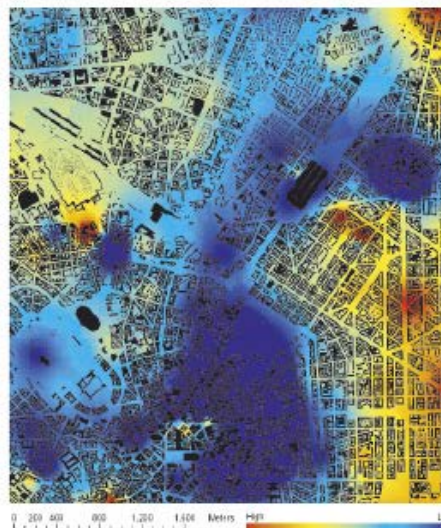


Últimos 30 días

Día 16 de octubre 2013



# DENSIDAD DE ACTIVIDAD DE TELÉFONOS MÓVILES EN EL ÁREA METROPOLITANA DE MILÁN ENTRE LAS 4 Y LAS 6 PM



# CONTEOS POR MEDIO DE CÁMARAS DE VIDEO Y CÁMARAS TÉRMICAS

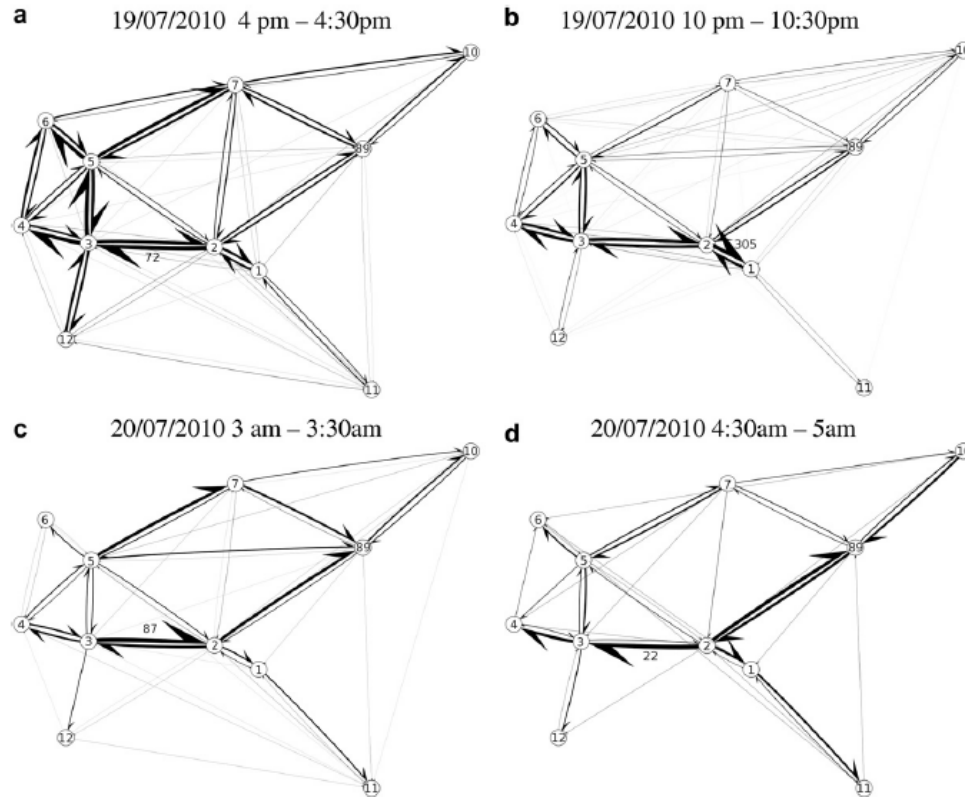
The proposed decision support system will analyze the video images from either visual cameras or thermal imaging cameras or both and automatically spot the problems of crowding and issues warning to the security staff to concentrate on particular area.

In ADCEM, people entering and going out of the doors will also be counted so that total estimate of people inside the mosque can be provided online to the security staff.

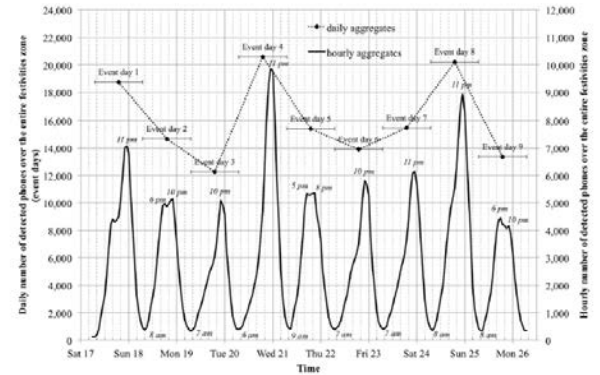
ADCEM estimates the information of crowd flow in the mosque that can help to observe the spots where crowd become stationary due to over-crowding.



# DETECTORES DE TELÉFONOS MÓVILES POR BLUETOOTH EN LA GESTIÓN DE MULTITUDES



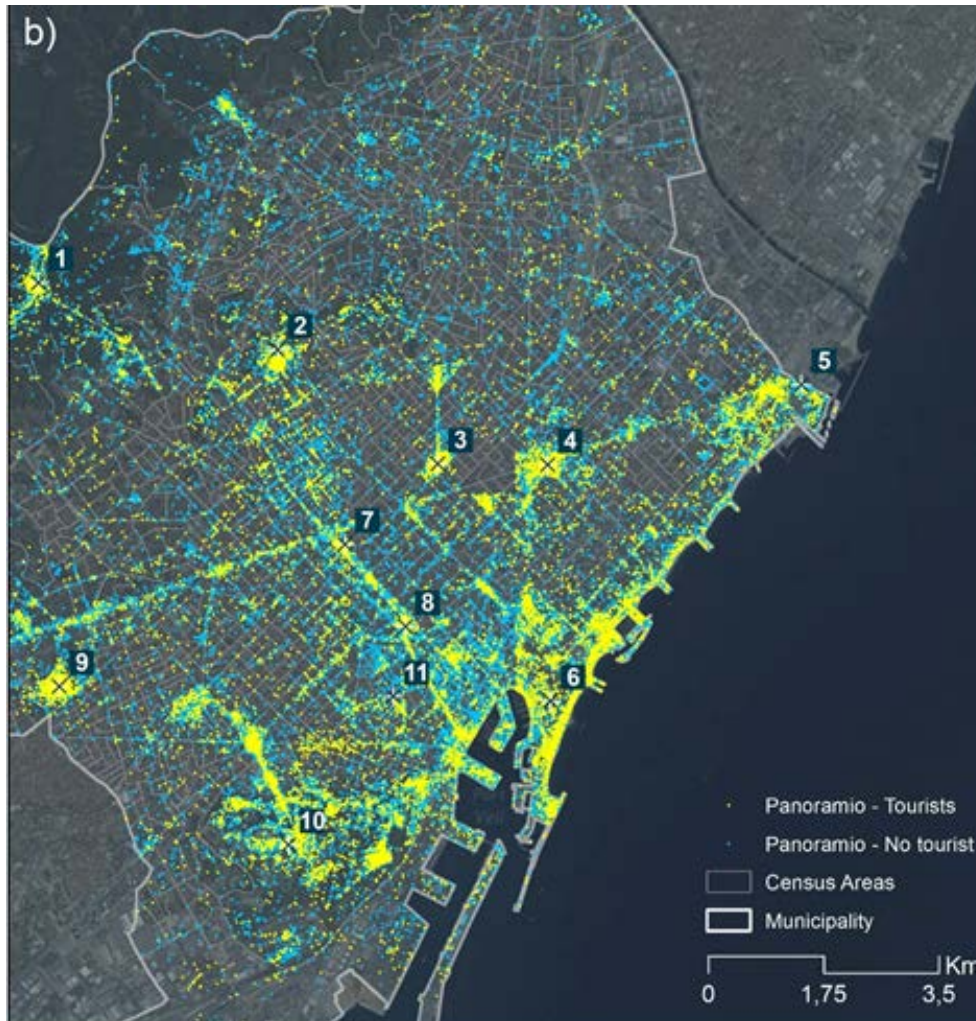
**Fig. 11.** Four snapshots of visitor flows in the festivities zone during periods of 30 min. The direction of the arrow indicates the direction of the flow, the width of the arrow indicates its size. The widths of the arrows are normalised to the size of the largest flow during each time period separately (indicated by the number next to the widest arrow). The numbers in the circles refer to Fig. 2. Locations 8 and 9 were generalised into one zone '89' for easier representation. Movements taking less than 30 s or longer than 30 min were discarded, and the remaining movements were allocated to the time period containing their departure time.



**Fig. 5.** Daily (event days starting and ending at 7 am) and hourly number of detected phones over the entire festivities zone as an indicator of crowdiness. Solid vertical gridlines point to midnight, dashed vertical gridlines are plotted every 4 h.

Versichele, M., Neutens, T., Delafontaine, M., & Van de Weghe, N. (2012). The use of Bluetooth for analysing spatiotemporal dynamics of human movement at mass events: A case study of the Ghent Festivities. *Applied Geography*, 32(2), 208-220.

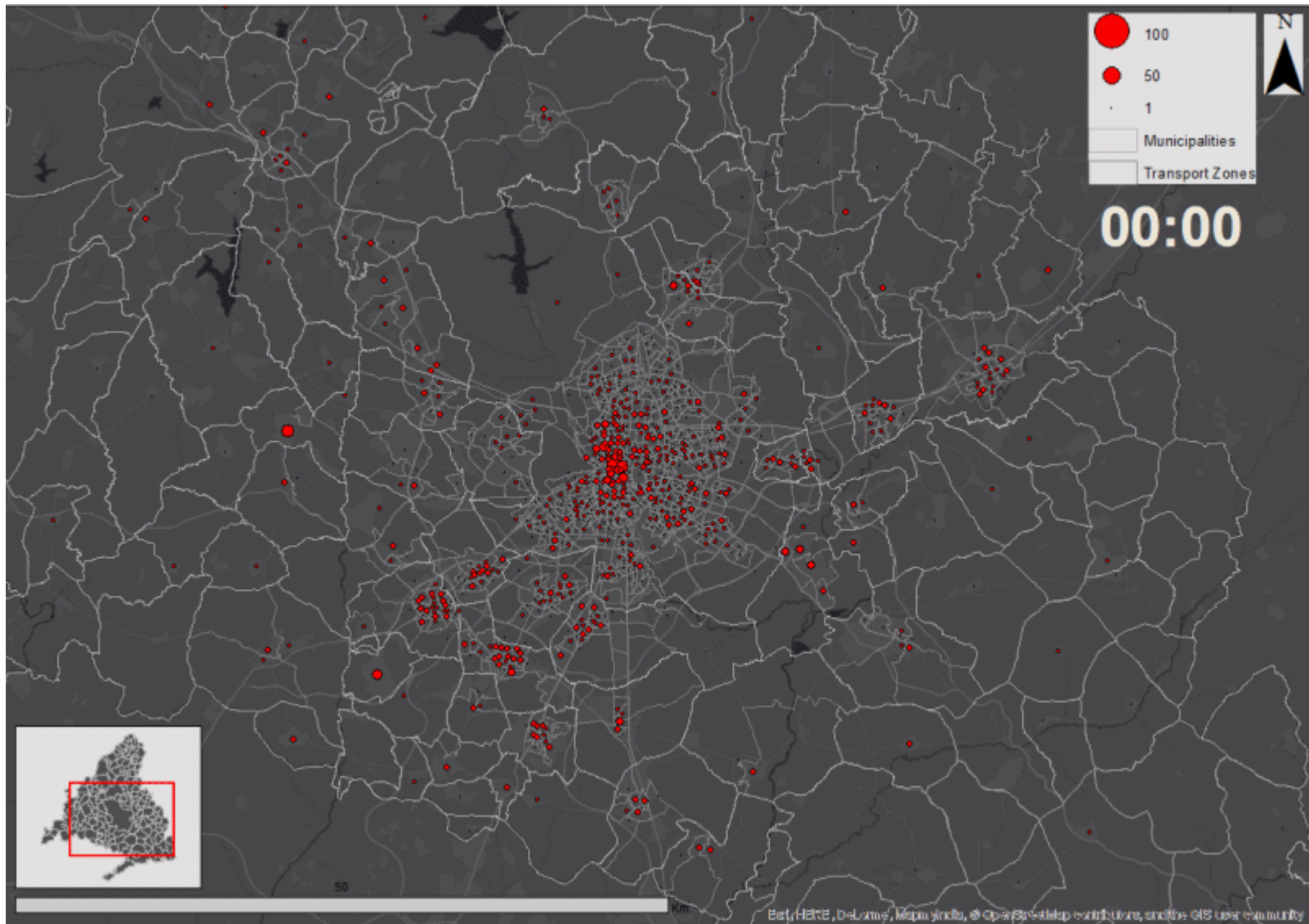
# LA HUELLA DIGITAL DE LOS TURISTAS. REDES SOCIALES DE FOTOGRAFÍAS GEOLOCALIZADAS: IDENTIFICANDO A LOS TURISTAS



Javier Gutiérrez, Juan Carlos García, María Henar Salas-Olmedo, Gustavo Romanillos. Airbnb irruption in heritage cities: towards the collapse of historical centres?



# EL PULSO DE LA CIUDAD A TRAVÉS DE LAS REDES SOCIALES: TWITTER



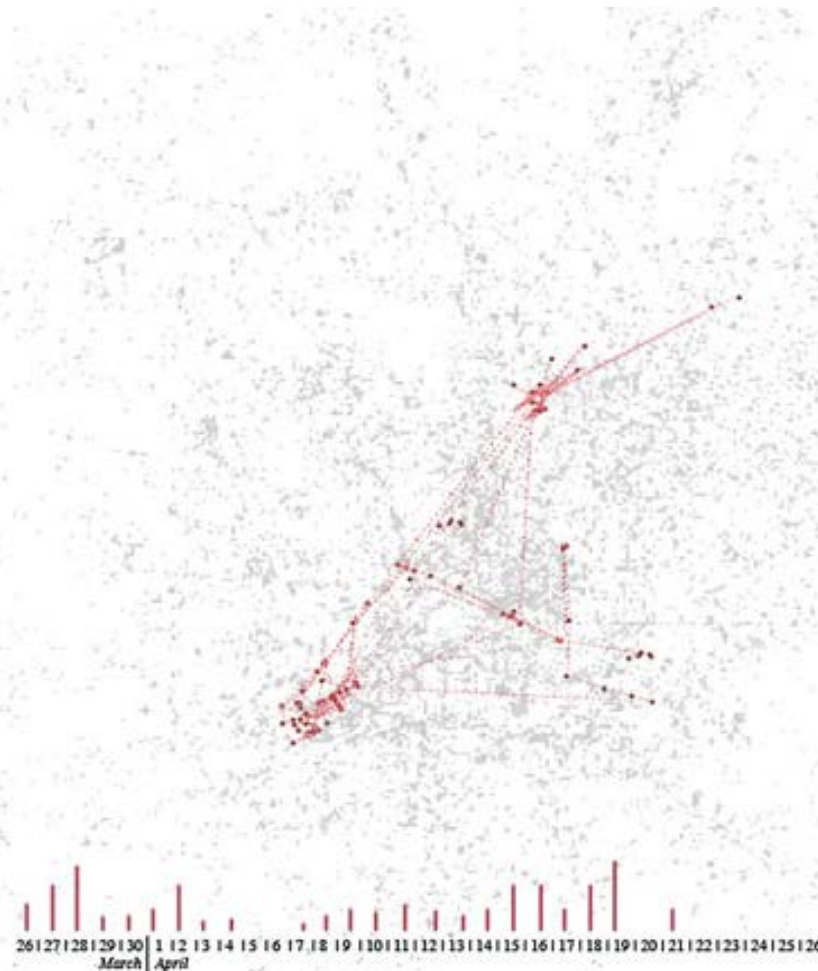
# TRAYECTORIAS PERSONALES: UN MES EN TWITTER

Telltale presents:

## *Urban Stories*

*...a month of Tweets in Milan*

@MartinaColombo  
@MaxMustache  
@MelissaFerrari  
@merist  
@mikomaki  
@muzzola  
@official\_sara  
@publin1  
@Bgs@jibela  
@LoZainetto  
@maecial  
@Marco\_GS  
@marcopetillo\_  
@MartinaColombo  
@MaxMustache  
@MelissaFerrari  
@merist  
@mikomaki  
@muzzola  
@official\_sara  
@publin1  
@Bgs@jibela  
@LoZainetto  
@maecial  
@Marco\_GS  
@marcopetillo\_  
@Mara29\_  
@MartinaColombo  
@MaxMustache  
@MelissaFerrari  
@merist  
@mikomaki  
@muzzola  
@official\_sara  
@publin1  
@Bgs@jibela  
@LoZainetto  
@maecial  
@Marco\_GS  
@marcopetillo\_



@MartinaColombo:

26.03.2013

@IA we've been working on the topic in the last few months, I will update you as soon as we release our viz :-)

27.03.2013

RT @giuseppebaldi: Really cool documentation of the workshop held by @aodebox team and Donato Ricci @densitydesign in Monza! Cal <http://t.co/4E>  
nice idea: <http://t.co/5WVGACJy6> Color signatures for classic novels based on the novel's visual subject matter (via @flowingdata)

27.03.2013

@merist\_itsfate: Thanks Merist! I am also exploring it on my PhD research <http://t.co/CwDp0Outf>! @densitydesign with @piccinardi :-)  
New Town Traces: notes of people passing through the New Towns throughout their day <http://t.co/bck2lRlj> on [UrbanSensing](http://t.co/UrbanSensing)

28.03.2013

RT @urbansensing: Next @UrbanSensing

26.03.2013

RT @albertocairo: Hello, Milan. Just landed. Talk tomorrow <http://t.co/Xa4kK456T>! Today, meeting with @piccinardi's students #dataviz #i4e

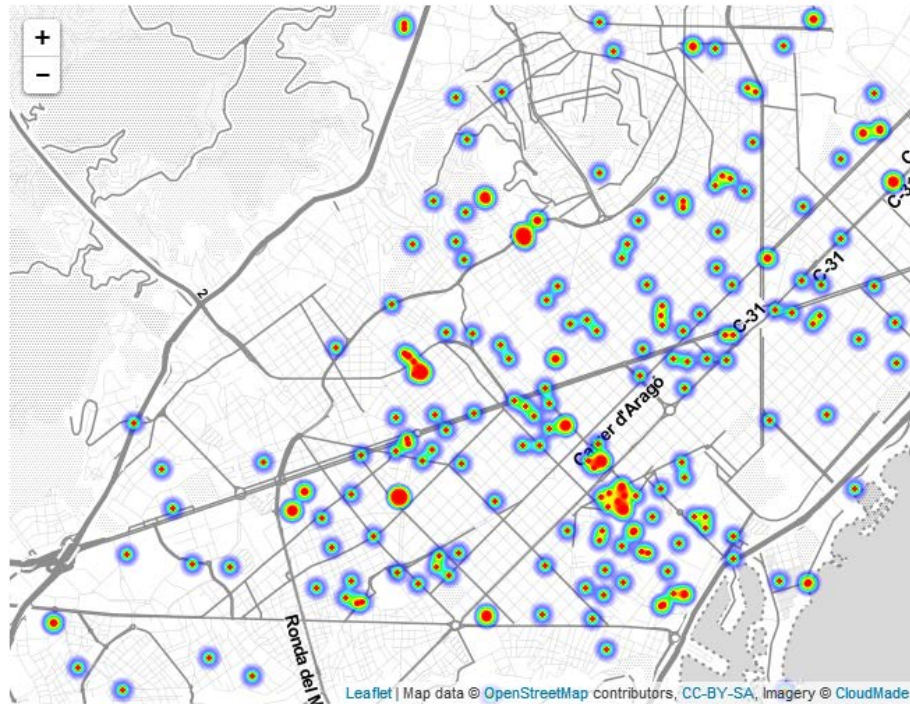
26.03.2013

RT @littlead: If you are in Milan tomorrow you can't miss @albertocairo talking about data visualization at Meet the Media Guru <http://t.co/4E>! Nice and elegant! <http://t.co/EYvFodqps4> The Evolution of Video Game Controllers, visualized (via @lowresolution)

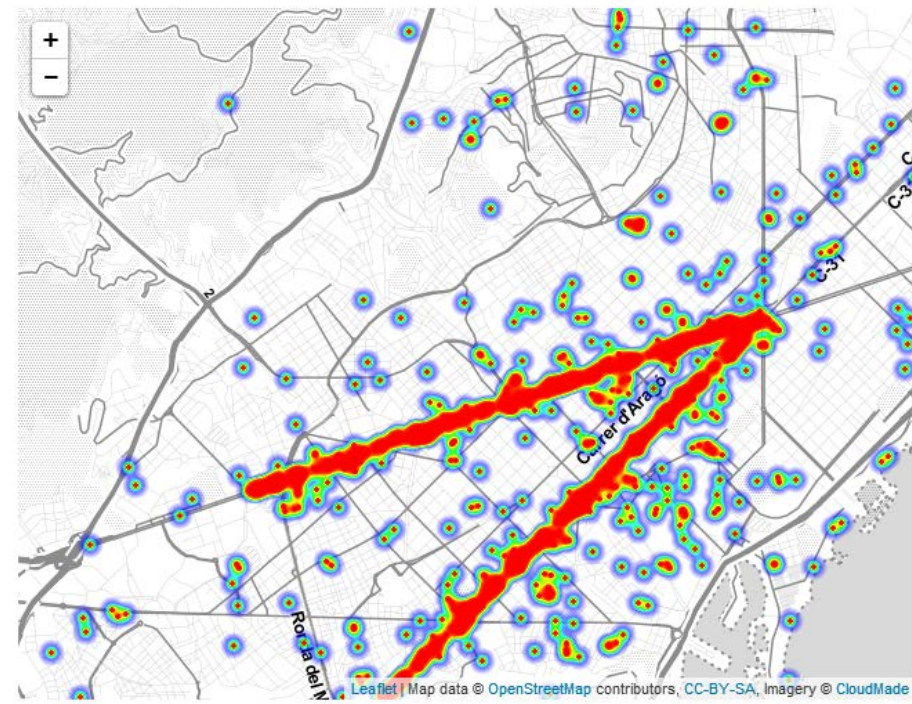
Ciuccarelli, P., Lupi, G., & Simeone, L. (2014): Visualizing the Data City (pp. 17-22). Springer International Publishing.

# IDENTIFICACIÓN DE EVENTOS ESPECIALES A PARTIR DE DATOS DE LAS REDES SOCIALES: TWITTER

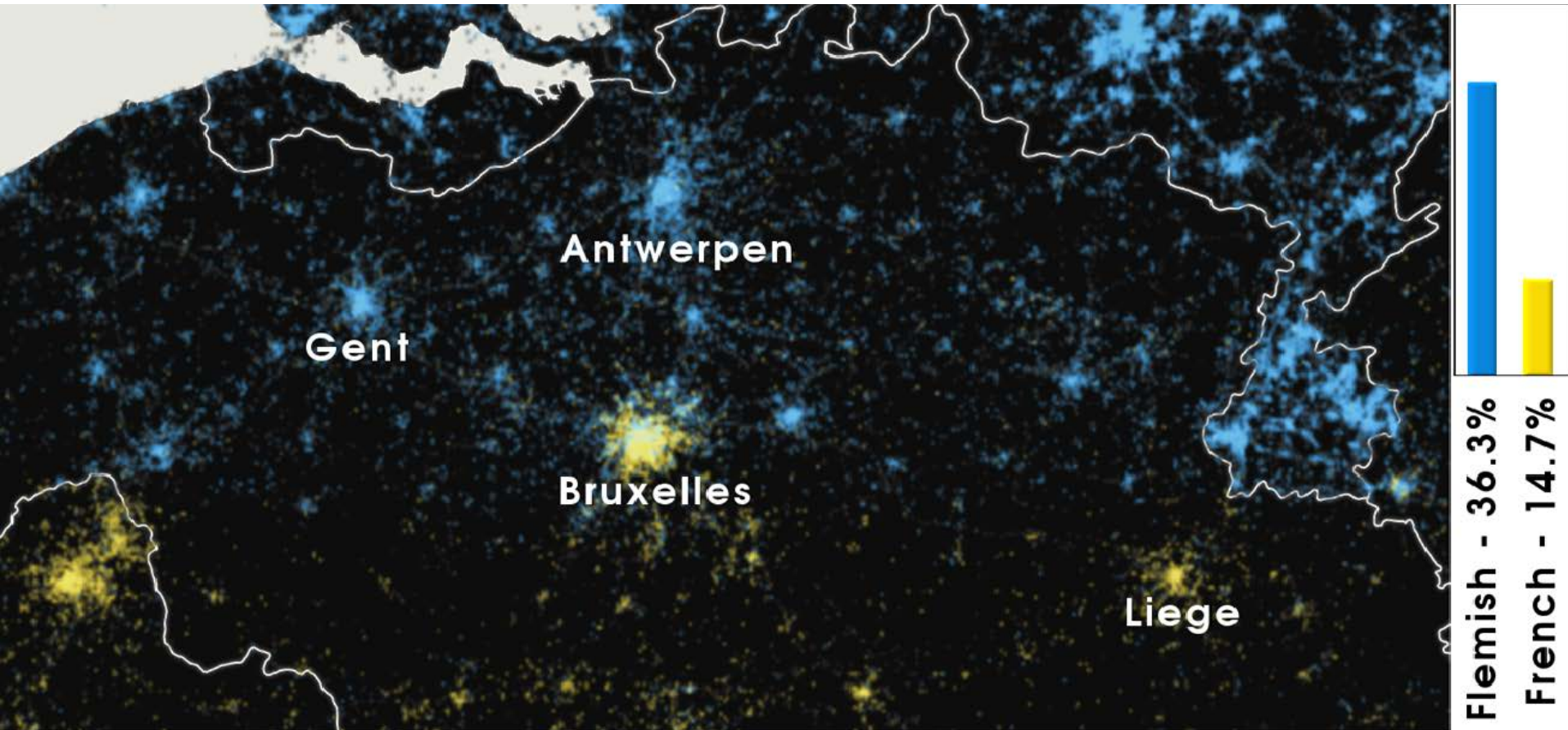
Tweets geolocalizados enviados en un día usual (10 de septiembre) entre las 16:00 y las 18:00h.



Tweets geolocalizados enviados hoy 11 de septiembre entre las 16:00 y las 18:00h.

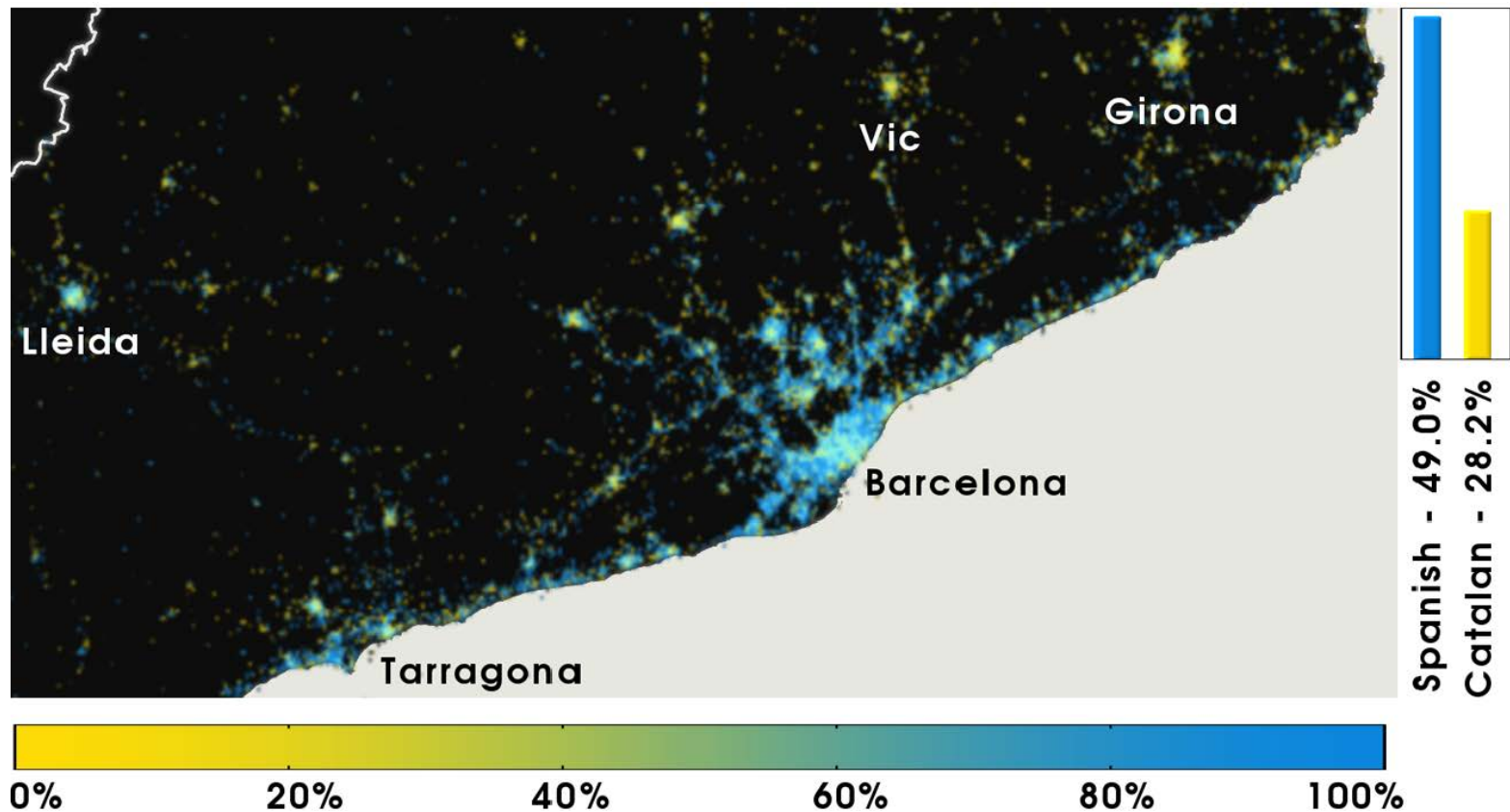


## LA LENGUA DE LOS TWEETS: FRONTERAS LINGÜÍSTICAS



Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q. y Vespignani, A. (2013). The twitter of babel: Mapping world languages through microblogging platforms. PLoS ONE, 8(4), e61981.

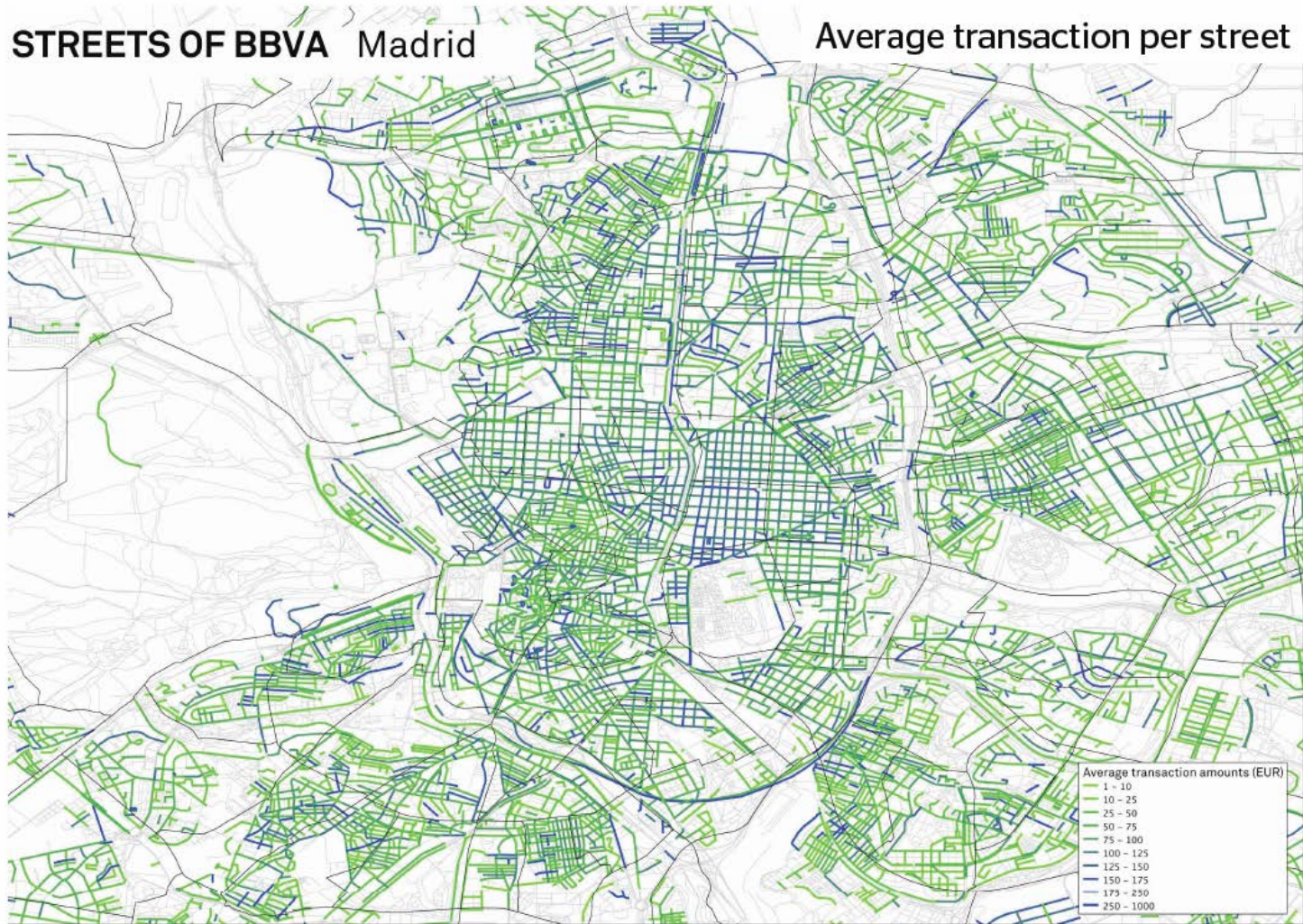
# LA LENGUA DE LOS TWEETS: BILIGÜISMO



Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q. y Vespignani, A. (2013). The twitter of babel: Mapping world languages through microblogging platforms. PLoS ONE, 8(4), e61981.

¿Cuánto?

## GASTO MEDIO CON TARJETA DE CRÉDITO EN MADRID



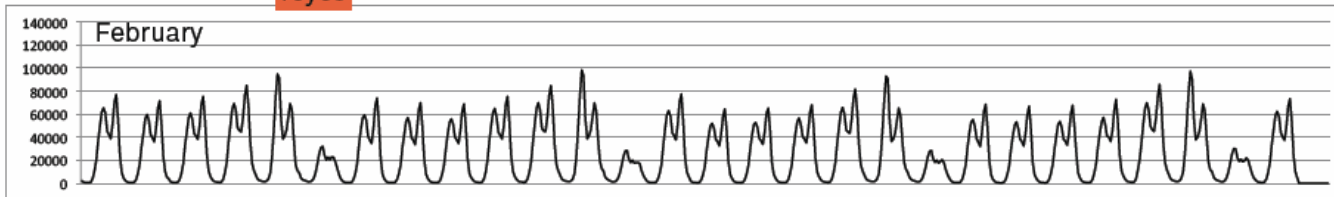
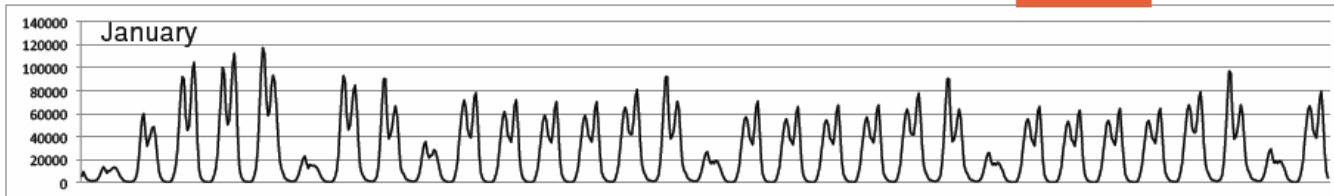
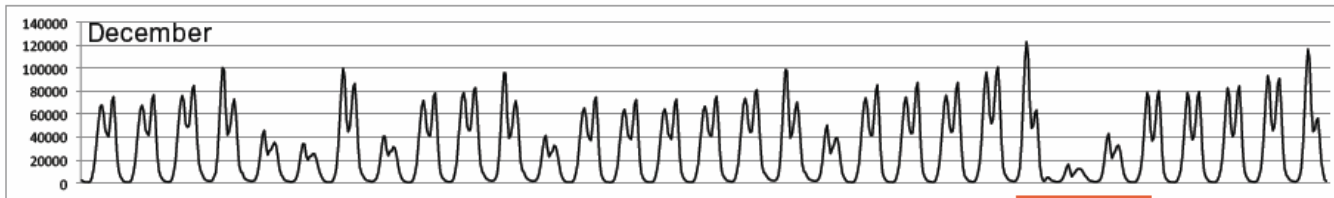
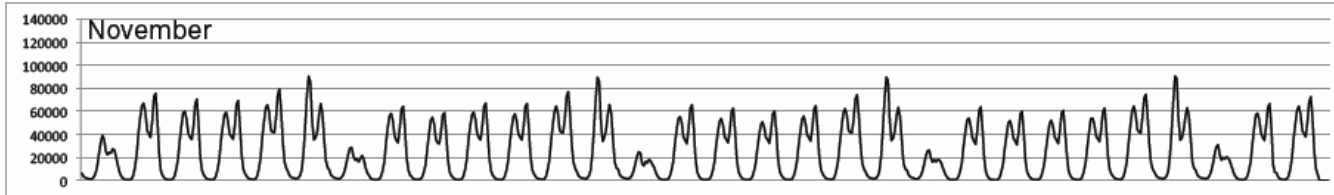
**BBVA**

<https://www.centrodeinnovacionbbva.com/noticias/6564-el-uso-eficiente-de-la-informacion-y-el-modelo-de-smart-cities-que-queremos>

¿Cuándo?

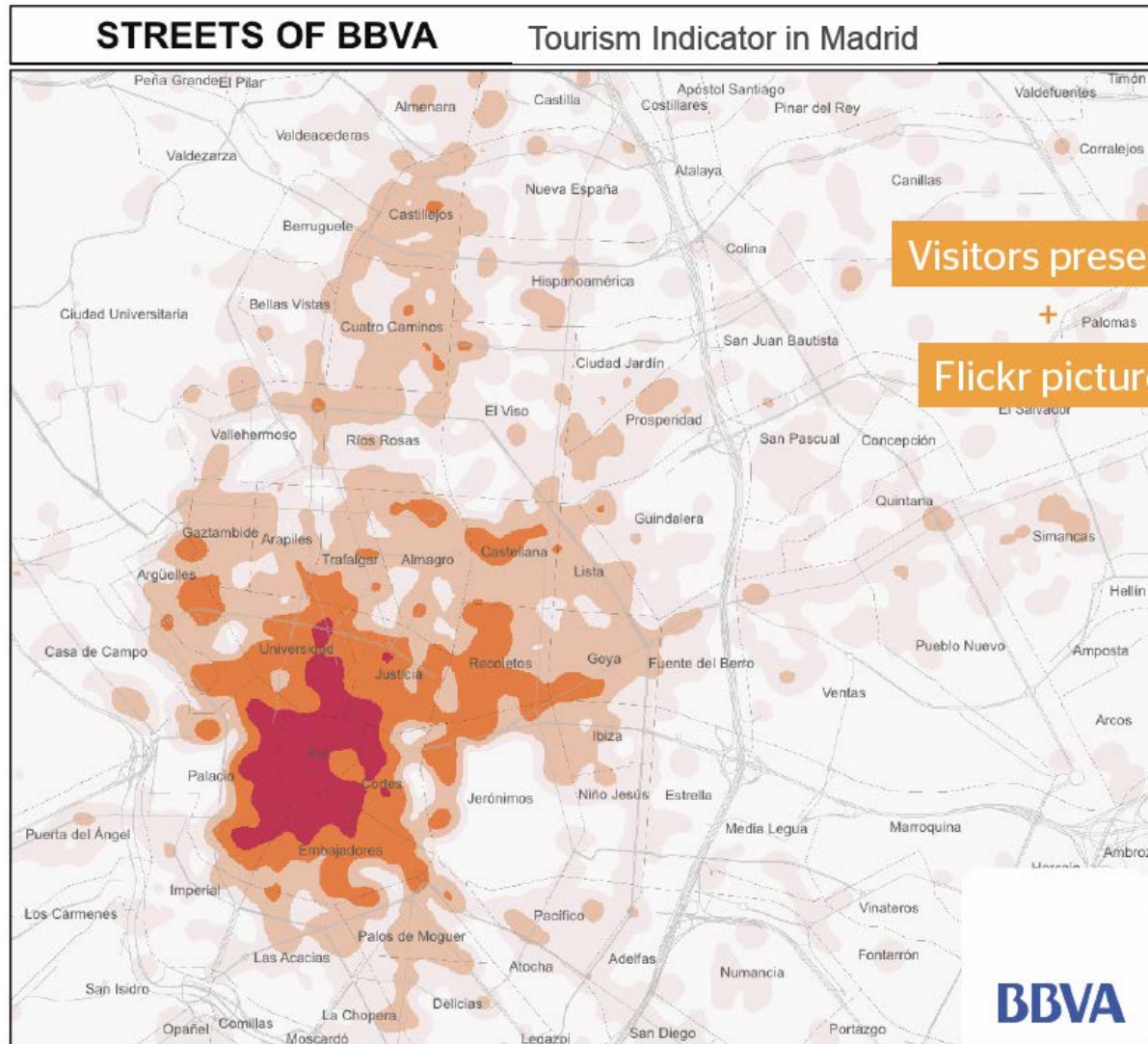
# PERFIL TEMPORAL DEL GASTO CON TARJETAS DE CRÉDITO EN MADRID

## Temporal signature of Madrid



¿Quién?

## LOCALIZACIÓN TURISTAS



<https://www.centrodeinnovacionbbva.com/noticias/6564-el-uso-eficiente-de-la-informacion-y-el-modelo-de-smart-cities-que-queremos>



## POSICIONES SOBRE LA REVOLUCIÓN DEL BIG DATA

- Es una panacea. Permite abordar todos los temas con mejores resultados que las fuentes tradicionales.
- No aporta información nueva. En el caso de las Ciencias Sociales, la información no es fiable.
- Aporta información complementaria a las fuentes tradicionales. Los datos deben ser depurados y contrastados. De los datos a la información y de la información al conocimiento

## ALGUNOS PUNTOS CRÍTICOS EN BIG DATA

- Dificultad de acceso a algunas fuentes de datos: dependencia de las empresas
- Dificultad del proceso de datos masivos y variados
- Sesgo / contraste con otras fuentes
- Privacidad/problemas legales
- Mal uso de la información



Tomado de Esteban Moro, Universidad Carlos III

# BIG DATA

Javier Gutiérrez Puebla  
Departamento de Geografía Humana  
Universidad Complutense de Madrid



tGIS

transporte infraestructuras y territorio  
Grupo de Investigación de la Universidad Complutense de Madrid