

# Ciencia de Redes (Humanas y Sociales)

Clase #6a

Carlos Sarraute

Grandata Labs

Abril - Junio 2019

# Agenda

- 1 Fuentes de Datos y Metodología
- 2 Flujo Urbano
- 3 Movilidad pendular
- 4 Movilidad y eventos sociales

# Objetivo de esta investigación

Usar registros de telefonía celular para bajar los costos de planeamiento urbano y de transporte.

Generamos una estimación de la distribución de la población entre diferentes partes de Buenos Aires usando CDR (Call Detail Records)

Nuestro objetivo: probar que estos métodos son confiables y extender el análisis a otras situaciones usando la masividad de los datos recolectados.

# Fuente de datos moviles

- Aplicamos nuestra metodología a la ciudad de Buenos Aires.
  - Centro político, financiero, cultural del país.
  - Tiene **2,890,151 habitantes**, divididos en **48 barrios** que se agrupan en **15 comunas**.
- Nuestro dataset de registros de llamados (CDRs)...
  - ...tiene llamados correspondientes a 4.95 millones de usuarios, en un periodo de 3 meses.
  - ...contiene, para cada llamado: origen, destinatario, hora, día, duración y antena (geolocalizada) usada para conectarse.
- Usamos esta información para asociar cada llamado a un barrio y una comuna.

# Matriz de Locaciones

Separamos una semana típica en 4 grupos de días y 4 grupos de horas:

<b>Grupos de días</b>	<b>Grupos de horas</b>	
Lunes a jueves	Mañana	5am - 11am
Viernes	Mediodía	11am - 3pm
Sabado	Tarde	3pm - 8pm
Domingo	Noche	8pm - 5am (dia sig.)

Esta selección de grupos de horas corresponde con el analisis basado en las encuestas de Origen - Destino ENMODO 2010, usada para determinar picos y valles de movilidad, para un día laboral tipico.

# Matriz de Locaciones

- Calculamos la proporción de llamados que un usuario  $u$  hace en la comuna  $c$  durante un grupo de días  $d$  y grupo de horas  $h$ :

$$P_{u,d,h,c} = \frac{R_{u,d,h,c}}{\sum_{c' \in \mathcal{C}} R_{u,d,h,c'}}$$

- La matriz  $P_u$  es lo que llamamos *Matriz de Distribución de Locaciones* del usuario  $u$ .

# Escalando con la población del censo

Determinamos la comuna correspondiente a la casa  $H_u$  para cada usuario  $u \in \mathcal{U}$ . Considerando que los usuarios están en sus casas los días de semana a la noche.

Con esta información, extendemos nuestras predicciones usando la información del censo INDEC (2010). Nos da un factor de escala  $F_c$  para cada comuna  $c$ :

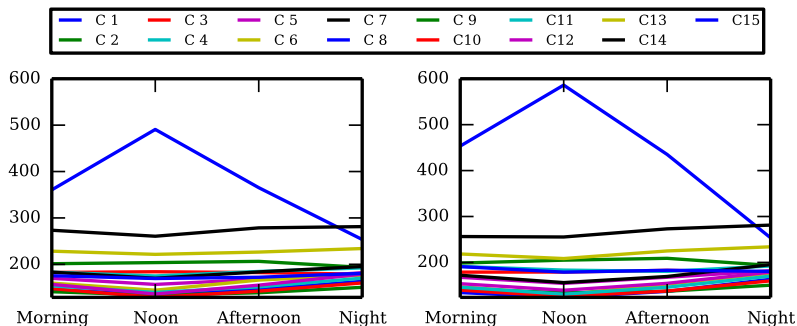
$$F_c = \frac{\text{pop}_c}{\#\{u \in \mathcal{U} | H_u = c\}}$$

# Agenda

- 1 Fuentes de Datos y Metodología
- 2 **Flujo Urbano**
- 3 Movilidad pendular
- 4 Movilidad y eventos sociales



# Validación del método



Comparación entre la encuesta ENMODO (izq.) y la metodología propuesta (der.)

# Validación del método

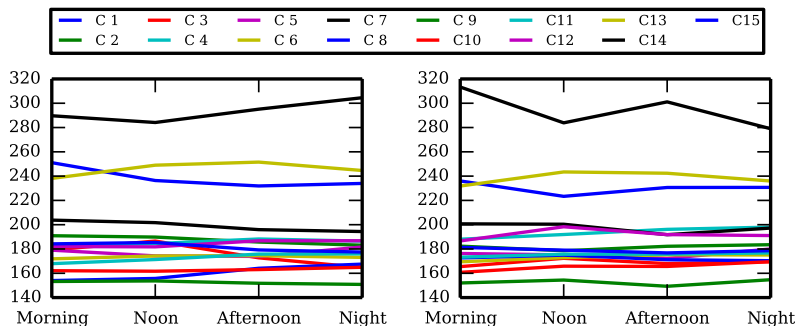
Validamos la metodología comparando con la encuesta de origen-destino ENMODO 2010, un método muy usado en planificación urbana.

Los resultados son similares (la diferencia promedio entre estimaciones es 5%).

Las mayores variaciones aparecen en la Comuna 1, en la mañana (20%) y la segunda en Commune 6, al mediodía (11%).

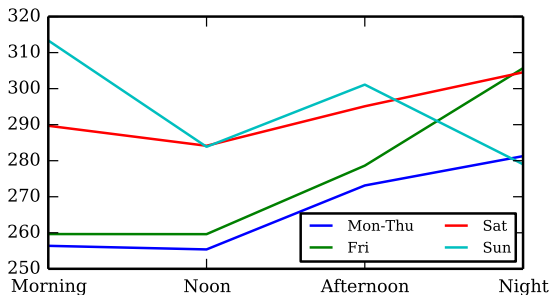
# Extensión a fines de semana

**Resultados nuevos** sobre movilidad (no cubiertos por las encuestas origen-destino).



Resultados para Sabado (izq) y Domingo (der).

## Analisis de Comuna 14 (Palermo)



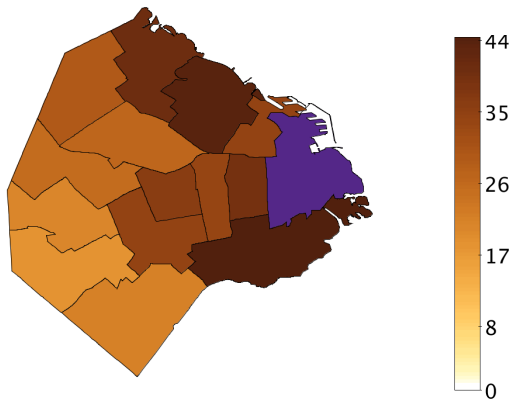
- Patrón residencial los días de semana.
- Polo de atracción los Sabados.
- Mayores picos en la noche del Viernes y Sabado.
- Domingo a la noche se parece más a la noche de un día de semana.

## Visualizando el pulso de la ciudad

Presentamos una comparación entre Comuna 1 y Comuna 6, para un día de semana al mediodía.

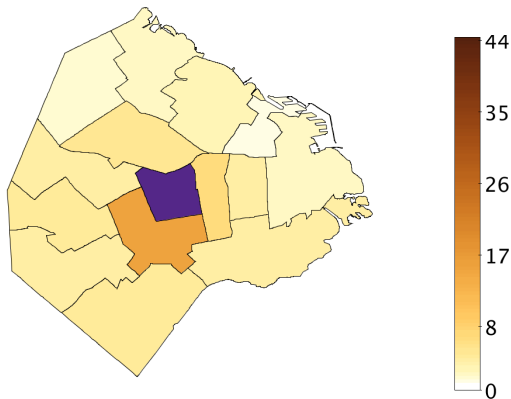
El gráfico muestra donde vive la gente que se encuentra en la comuna estudiada (marcada en violeta) en ese horario. La escala está en miles de personas.

## Pulso de ciudad - Comuna 1 (Microcentro)



- Población moderada.
- Distrito financiero y comercial.
- Atrae gente incluso de zonas distantes.

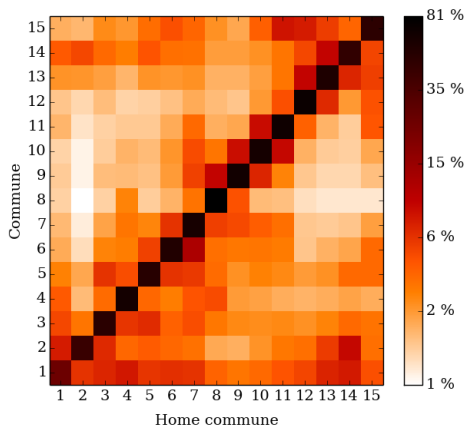
## Pulso de ciudad - Comuna 6 (Caballito)



- Una de las areas mas densamente pobladas.
- Zona mayormente residencial.
- Atrae menos gente, y de barrios vecinos.

# La matriz del pulso de la ciudad

Con la información de movilidad también podemos generar la *matriz del flujo* entre comunas, acá para un día típico al mediodía:





# Agenda

- 1 Fuentes de Datos y Metodología
- 2 Flujo Urbano
- 3 Movilidad pendular**
- 4 Movilidad y eventos sociales

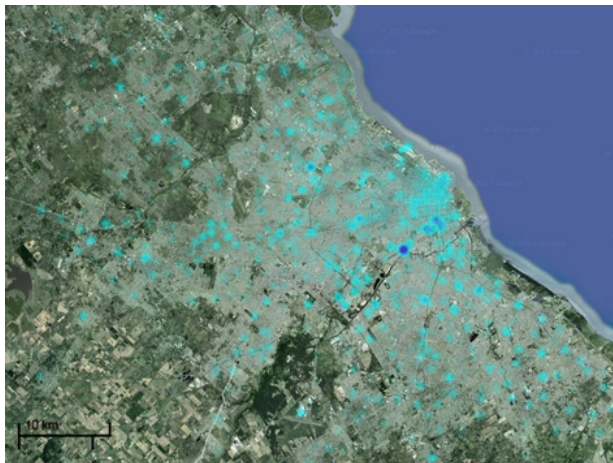
# Movilidad pendular del GBA a la CABA

- Los viajes pendulares (casa-trabajo) son un fenomeno prevalente en grandes areas urbanas, y aparecen naturalmente en datos de telefonía movil.
- En este caso estudiamos gente que:
  - Vive en GBA (Gran Buenos Aires), donde los vemos a la noche.
  - Se encuentra en CABA (Capital Federal) durante el día.
- De los datos, estimamos la distancia recorrida entre casa y trabajo (*radius of commute*)
  - Consideramos como casa y trabajo las antenas mas usadas. Csáji et al. (2012)
  - Obtuvimos un radio de viaje promedio de 7.8 km
  - Como comparación, el diametro de la ciudad es de 14 km, y el diametro del area metropolitana es 90 km.

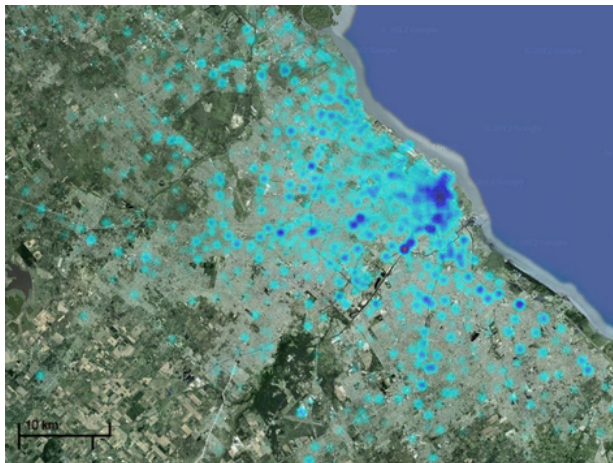
# Movilidad entre GBA y CABA - 05:00



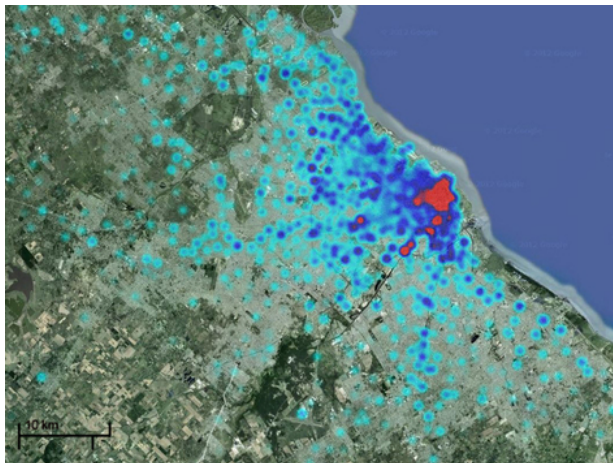
# Movilidad entre GBA y CABA - 06:00



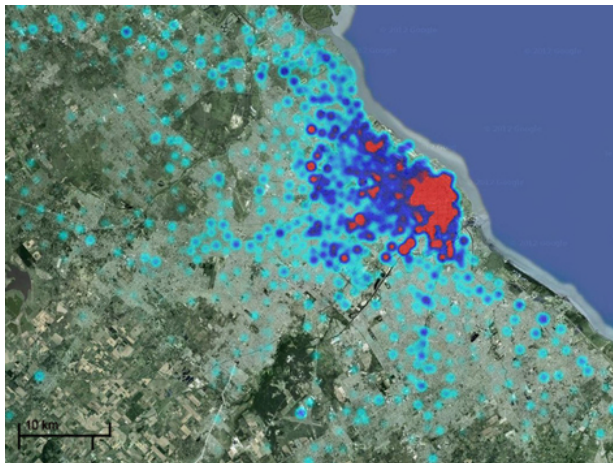
# Movilidad entre GBA y CABA - 07:00



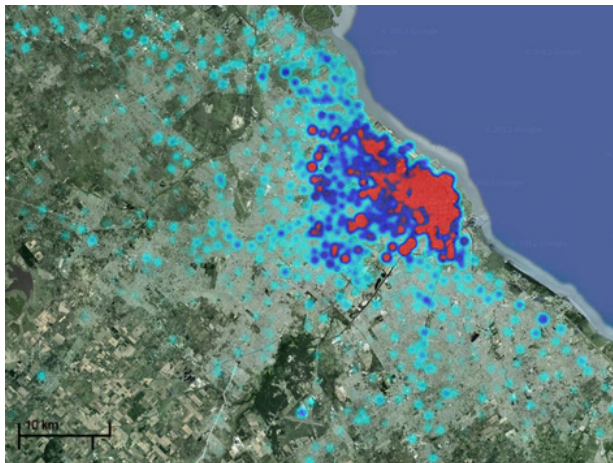
# Movilidad entre GBA y CABA - 08:00



# Movilidad entre GBA y CABA - 09:00

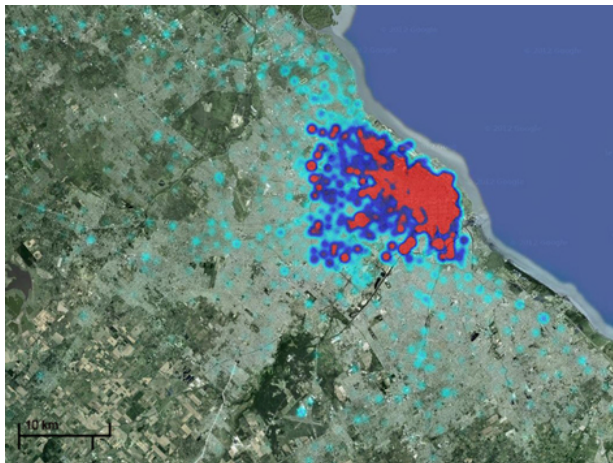


# Movilidad entre GBA y CABA - 10:00

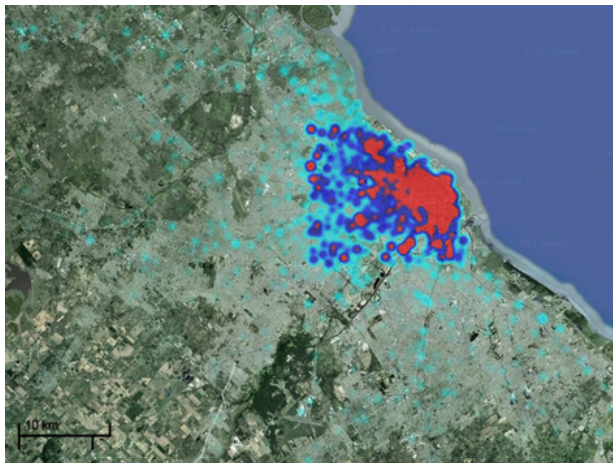




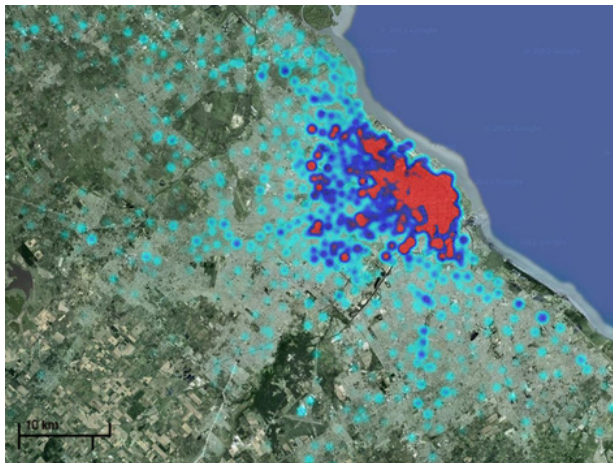
# Movilidad entre GBA y CABA - 12:00



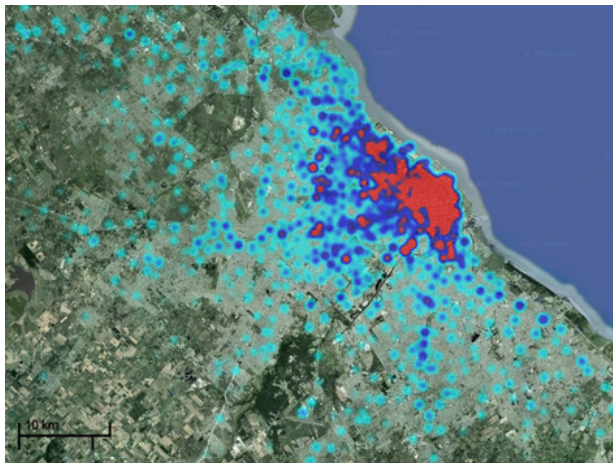
# Movilidad entre GBA y CABA - 14:00



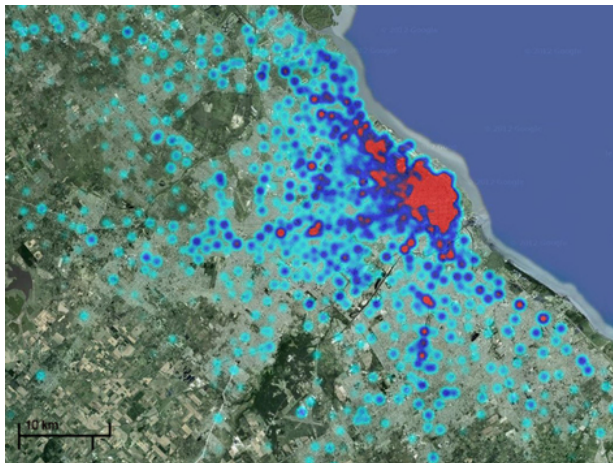
# Movilidad entre GBA y CABA - 16:00



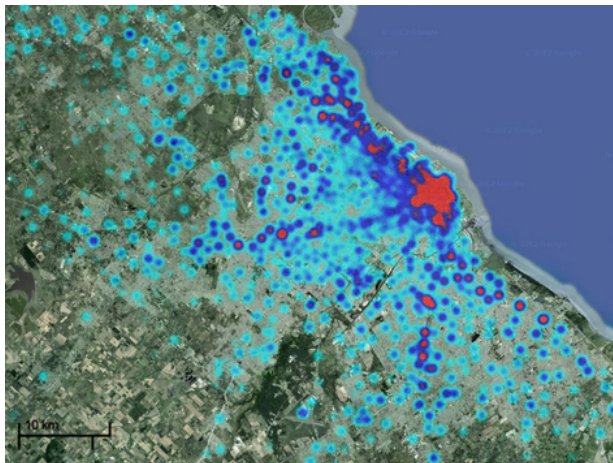
# Movilidad entre GBA y CABA - 17:00



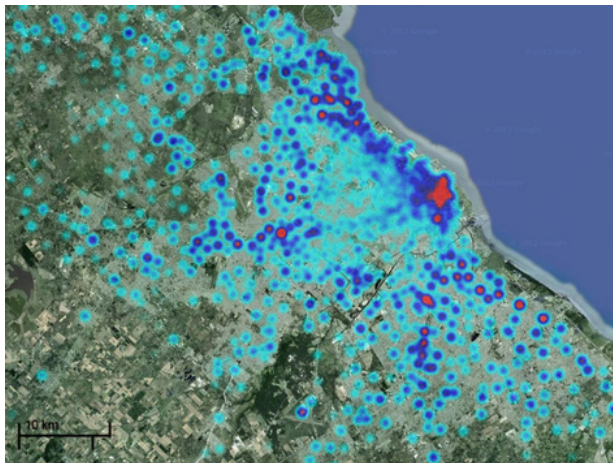
# Movilidad entre GBA y CABA - 18:00



# Movilidad entre GBA y CABA - 19:00



# Movilidad entre GBA y CABA - 20:00

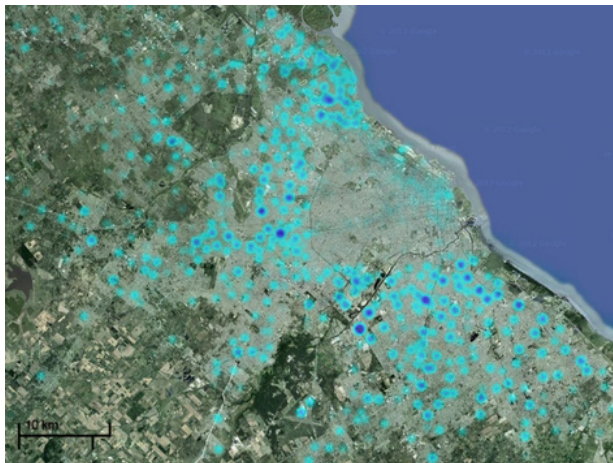


# Movilidad entre GBA y CABA - 21:00





# Movilidad entre GBA y CABA - 23:00



# Agenda

- 1 Fuentes de Datos y Metodología
- 2 Flujo Urbano
- 3 Movilidad pendular
- 4 Movilidad y eventos sociales**

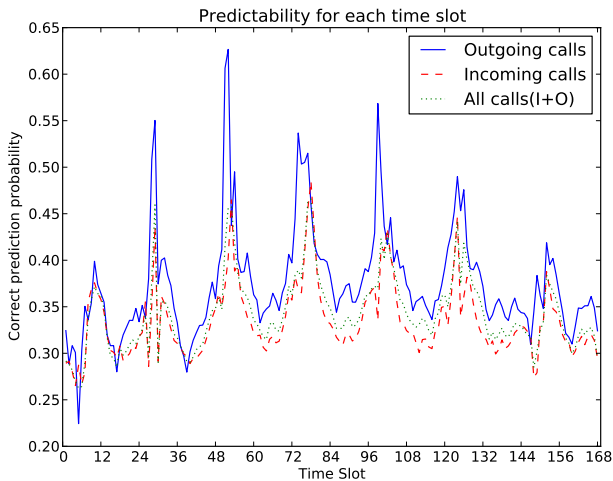
# Regularity of mobility

We split the week in time slots, one for each hour, totalizing  $7 * 24 = 168$  slots per week.

Humans tend to have very predictable mobility patterns Song et al. (2010); Gonzalez et al. (2008); Jiang et al. (2012). This model was used as a baseline in Cho et al. (2011).

Good predictability baseline, achieving an average of around **35%** correct predictions for a period of 2 weeks, training with 15 weeks of data, including peaks of over **50%** predictability.

# Basic prediction



# Event Index

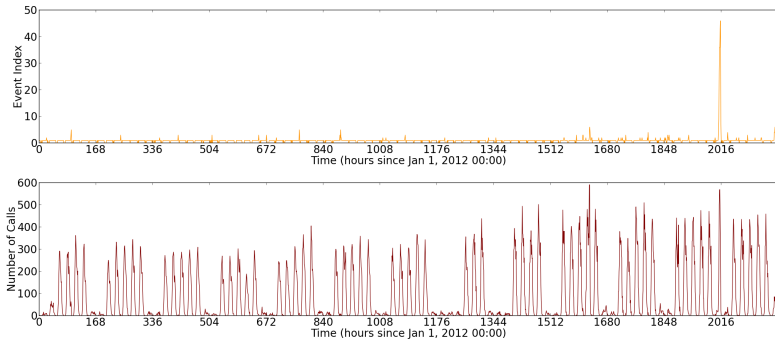
We consider the number of calls per hour registered in each antenna, according to:

- week  $w_i$  ( $0 \leq i < n$ , where  $n$  is the number of weeks in the studied dataset),
- day of the week  $d_j$  ( $0 \leq j \leq 6$ )
- hour of the day  $h_k$  ( $0 \leq k \leq 23$ ).

For each time slot  $(w_i, d_j, h_k)$ , we denote the number of calls as  $C_{(i,j,k)}$  and we compute an *event index*  $E_{(i,j,k)}$  given by the ratio

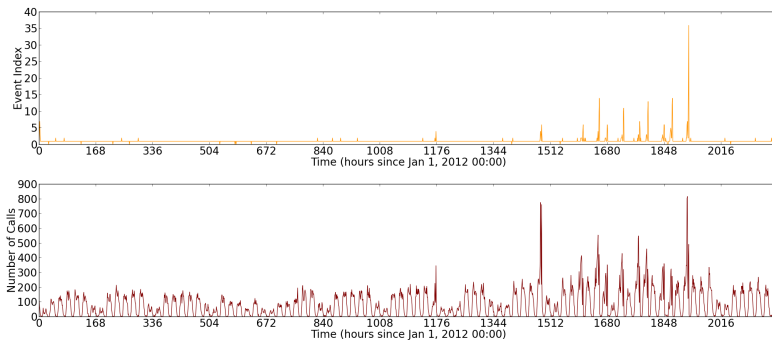
$$E_{(i,j,k)} = \frac{C_{(i,j,k)}}{\frac{1}{n-1} \sum_{0 \leq \ell < n, \ell \neq i} C_{(\ell,j,k)}}. \quad (1)$$

# Massive event (microcentro)



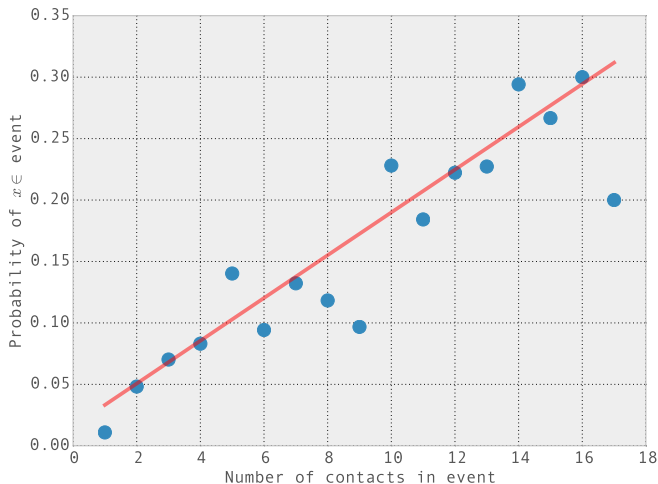
**Figure:** Event index  $E$  and volume of calls registered in a cell tower near Plaza de Mayo for each time slot (hour) during a period of 3 months.

# Massive events (estadio River)



**Figure:** Event index  $E$  and volume of calls registered in a cell tower near River Plate Stadium for each time slot (hour) during a period of 3 months. Several massive events were detected during March 2012 (the Roger Waters concerts).

# Probability of being in event



**Figure:** Conditional probability of a user being in an event given that  $n_c$  of his/her contact were present in the event.



# Combining Places, Transitions and Contacts

- “MobHet: Predicting human mobility using heterogeneous data sources”
- Lucas M. Silveira, Jussara M. de Almeida, Humberto T. Marques-Neto, Carlos Sarraute, Artur Ziviani
- Elsevier Computer Communications: Volume 95, 1 December 2016, Pages 54-68
- <https://doi.org/10.1016/j.comcom.2016.04.013>

# MobHet: Predicting human mobility

P = Popularity of regions

T = Probability of Transitions between regions

C = Contacts

		MobilePhone-MX Dataset				
Contact definition		<i>MobHet-C</i>	<i>MobHet-TC</i>	<i>MobHet-PC</i>	<i>MobHet-TPC</i>	<i>MobHet-TP</i>
<i>Contact Strength</i>	$\theta = 10\%$	$0.493 \pm 0.0232$	$0.593 \pm 0.0191$	$0.613 \pm 0.0193$	$0.630 \pm 0.0201$	
	$\theta = 15\%$	$0.509 \pm 0.0222$	$0.602 \pm 0.0195$	$0.626 \pm 0.0182$	$0.649 \pm 0.0213$	
	$\theta = 20\%$	$0.524 \pm 0.0214$	$0.623 \pm 0.0180$	$0.642 \pm 0.0190$	$0.668 \pm 0.0210$	
	$\theta = 25\%$	<b><math>0.544 \pm 0.0230</math></b>	<b><math>0.641 \pm 0.0198</math></b>	<b><math>0.663 \pm 0.0198</math></b>	<b><math>0.692 \pm 0.0172^*</math></b>	
	$\theta = 50\%$	<b><math>0.532 \pm 0.0218</math></b>	<b><math>0.638 \pm 0.0208</math></b>	<b><math>0.652 \pm 0.0203</math></b>	<b><math>0.683 \pm 0.0168^*</math></b>	
$\theta = 75\%$		$0.468 \pm 0.0212$	$0.575 \pm 0.0212$	$0.583 \pm 0.0215$	$0.618 \pm 0.0178$	
None						<b><math>0.619 \pm 0.0153</math></b>

# Referencias

- Secretaría de Transporte. Ministerio del Interior y Transporte. ENMODO (2009-2010). Resultados de la encuesta origen destino. Movilidad en el area metropolitana de Buenos Aires, 2010.
- Instituto Nacional de Estadística y Censos (INDEC). *Censo Nacional de Población, Hogares y Viviendas 2010*, volume 1. INDEC, October 2010.
- Balázs Cs Csáji, Arnaud Browet, VA Traag, Jean-Charles Delvenne, Etienne Huens, Paul Van Dooren, Zbigniew Smoreda, and Vincent D Blondel. Exploring the mobility of mobile phone users. *Physica A: Statistical Mechanics and its Applications*, 2012.
- Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.
- Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *Nature*, 453(7196):779–782, 2008.
- Shan Jiang, Joseph Ferreira, and Marta C González. Clustering daily patterns of human activities in the city. *Data Mining and Knowledge Discovery*, pages 1–33, 2012.
- Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *ACM SIGKDD*, pages 1082–1090. ACM, 2011.