



Afi Escuela
de Finanzas

Redes Sociales

Máster Executive en Data Science y Big Data

Alejandro Llorente Pinto

2022-05-03

Índice

1. Contagio social
2. Homofilia en redes sociales
3. Paradoja de la amistad
4. Enlaces débiles en redes sociales
5. Ley de Gravedad
6. Modelos de difusión en redes sociales

Conceptos avanzados en redes sociales

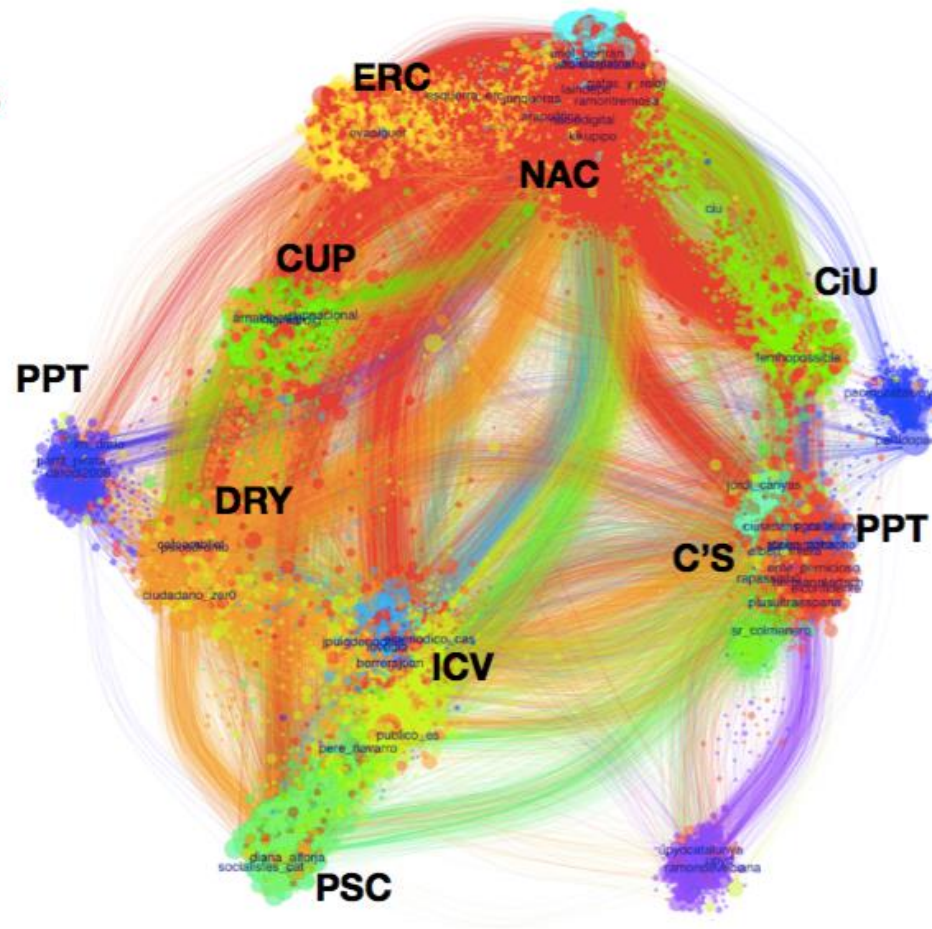
-
7. Redes sociales en política
 8. Redes sociales en marketing viral
 9. Redes sociales para detección de fraude
 10. Redes sociales en finanzas

Aplicaciones en redes sociales

1 | Conceptos avanzados en redes sociales

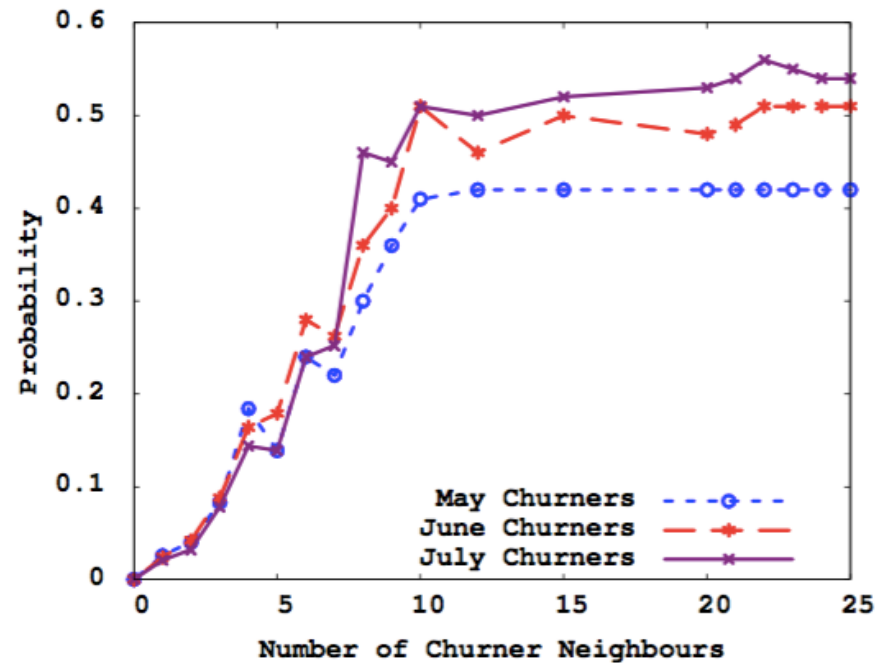
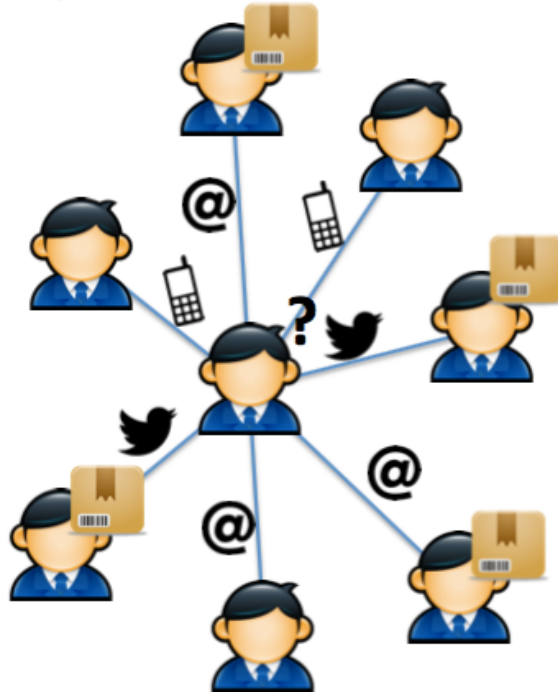
Homofilia en redes sociales

- En las redes sociales tendemos a relacionarnos con gente similar “Birds feather flock together”
- Personas relacionadas socialmente tienden a compartir gustos, lugar de residencia, opinión política, etc.
- La asortatividad de las redes es “homofilia de grado”



Contagio social

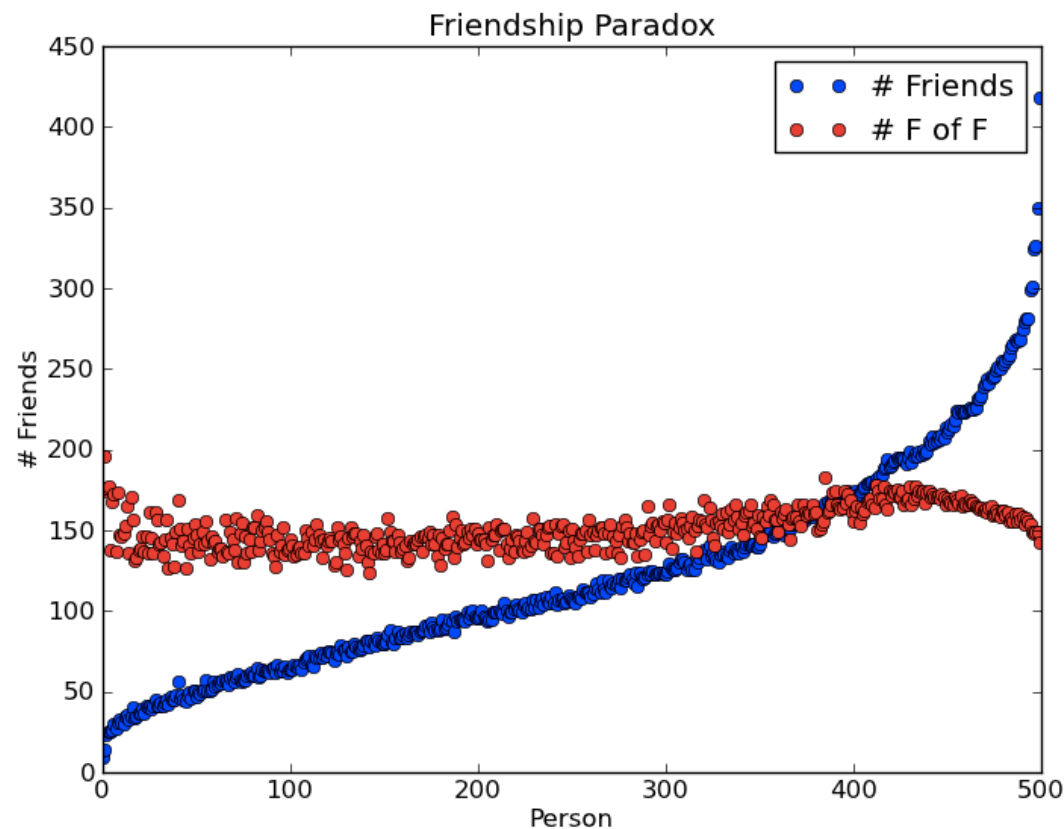
- Los comportamientos en la red son “contagiosos”
- Ejemplos:
 - Salud
 - Transmisión de información
 - Compra de productos, etc.



Dasgupta, K. et al., 2008. Social ties and their relevance to churn in mobile telecom networks.

Paradoja de la amistad

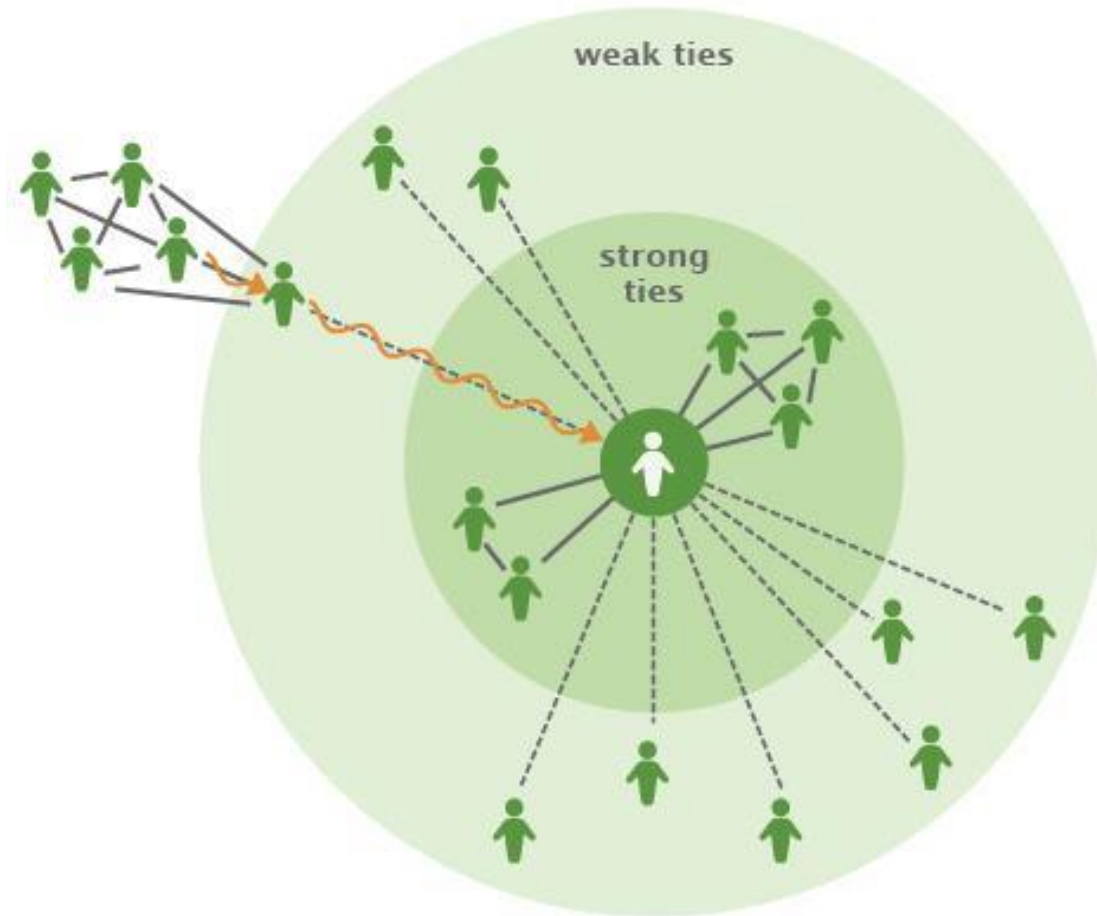
- ¿Pensáis que tenéis más o menos amigos que vuestros amigos?



Enlaces débiles

- Mark Granovetter, *La fuerza de los enlaces débiles*, 1973.
- En una red, podemos diferenciar aquellos enlaces fuertes, donde se observan interacciones habituales entre contactos y que siguen leyes de homofilia, comunidades, etc.
- Pero todos tenemos contactos con los que hablamos muy de vez en cuando, con los que hemos dejado de hablar, contactos que no sabríamos encajar en un grupo de amigos ni de trabajo, ... Estos son los llamados **enlaces débiles**.

Enlaces débiles



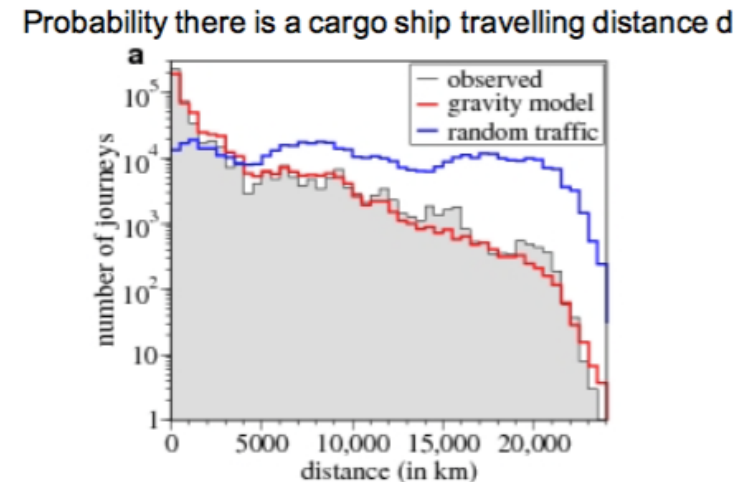
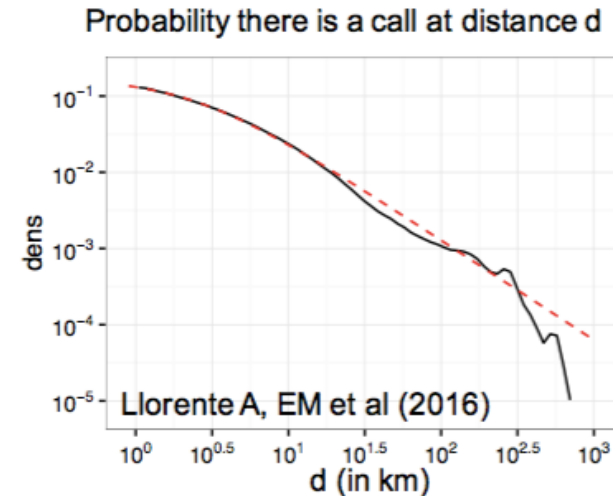
- Contactos que no siguen propiedades de clustering: enlaces entre dos miembros pero no entre sus comunidades.
- Muy relacionados con enlaces y nodos puente entre comunidades.
- Granovetter demostró que estos enlaces influyen de manera muy fuerte en procesos sociales como la búsqueda de trabajo.

Ley de Gravedad

- Las redes están embebidas en la geografía
- En las redes sociales hay una “ley de la gravedad”: la probabilidad de que haya un enlace a distancia d decae con la distancia

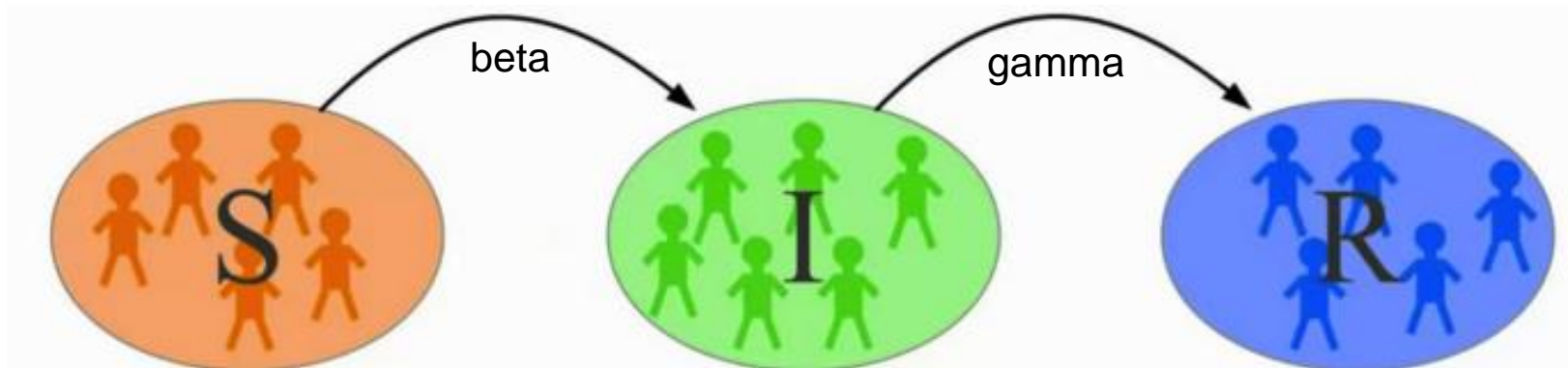
$$P(d) \sim d^{-\gamma}$$

- Es decir:
 - La mayoría de nuestros amigos están cerca
 - Pero algunos están lejos
- Las redes de transporte también tienen ley de la gravedad
- Las redes tecnológicas suelen decaer más rápido



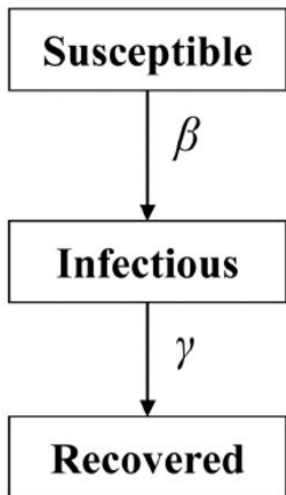
Modelos de difusión de información

- Existen diferentes modelos de difusión de información en redes sociales pero, esencialmente, todos son variaciones del modelos SIR:
 - (S)usceptible: población que puede contagiarse
 - (I)nfectado: población que está infectada
 - (R)ecovered: población que estuvo infectada, se recuperó y ya no puede contagiarse de nuevo

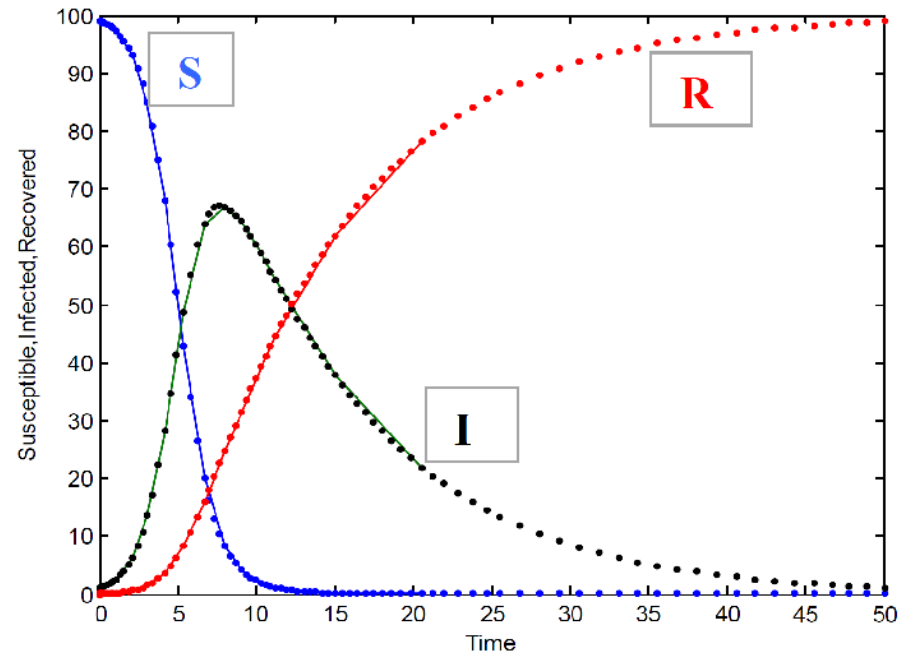


Modelos de difusión de información

- Estos modelos se pueden simular mediante ecuaciones diferenciales de una forma muy sencilla:

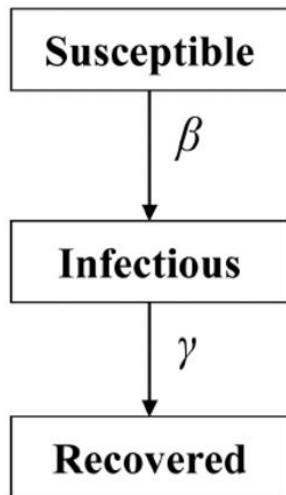


$$\begin{aligned}\frac{dS}{dt} &= -\beta SI \\ \frac{dI}{dt} &= \beta SI - \gamma I \\ \frac{dR}{dt} &= \gamma I\end{aligned}$$



Modelos de difusión de información

- Estos modelos se pueden simular mediante ecuaciones diferenciales de una forma muy sencilla:



$$\frac{dS}{dt} = -\beta SI$$

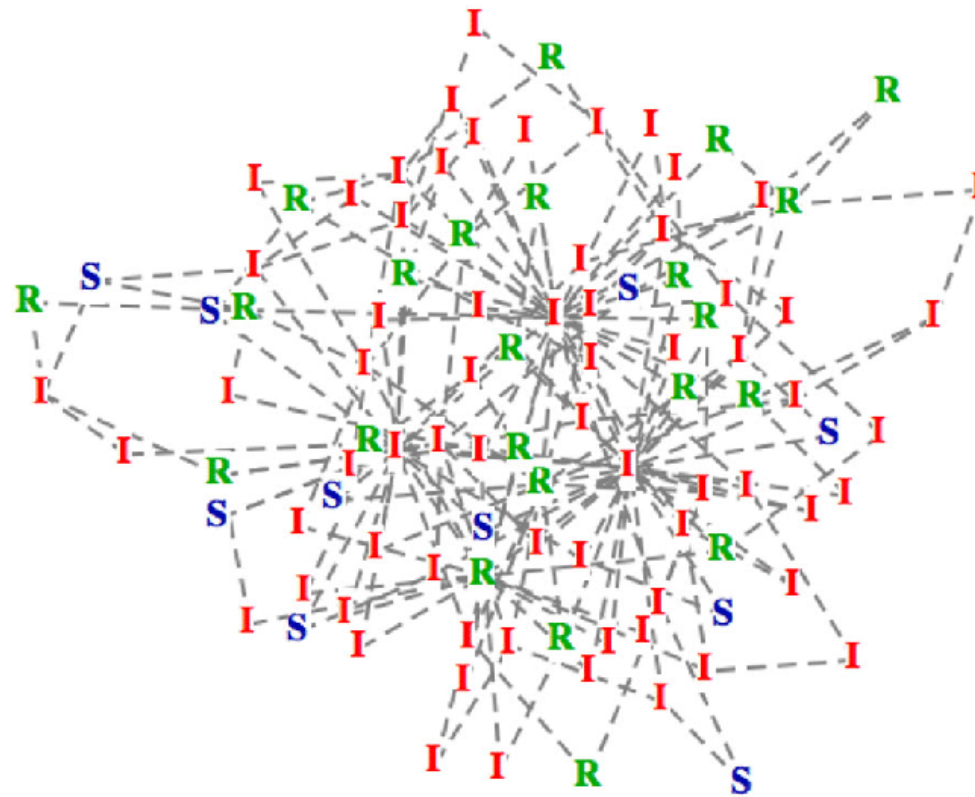
$$\frac{dI}{dt} = \beta SI - \gamma I$$

$$\frac{dR}{dt} = \gamma I$$

- Por unidad de tiempo, el número de personas susceptibles proporcionalmente a la probabilidad de transmisión y al producto del número de susceptibles por el número de infectados.
- Por unidad de tiempo, el número de personas infectadas aumenta proporcionalmente a la probabilidad de transmisión y al producto del número de susceptibles por el número de infectados y disminuye de acuerdo a la probabilidad de recuperación.
- Por unidad de tiempo, el número de personas recuperadas aumenta proporcionalmente al número de infectadas.

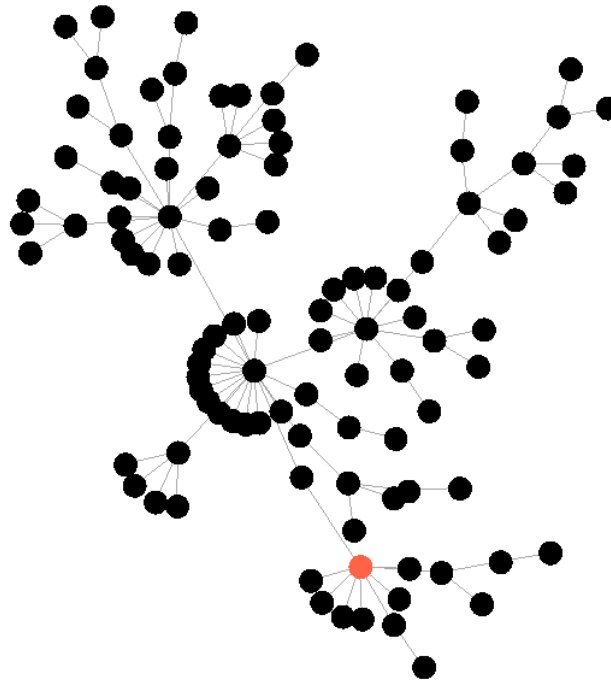
Modelos de difusión de información

- Pero también con redes sociales (último ejercicio de la práctica):



Modelos de difusión de información

- Pero también con redes sociales (último ejercicio de la práctica):

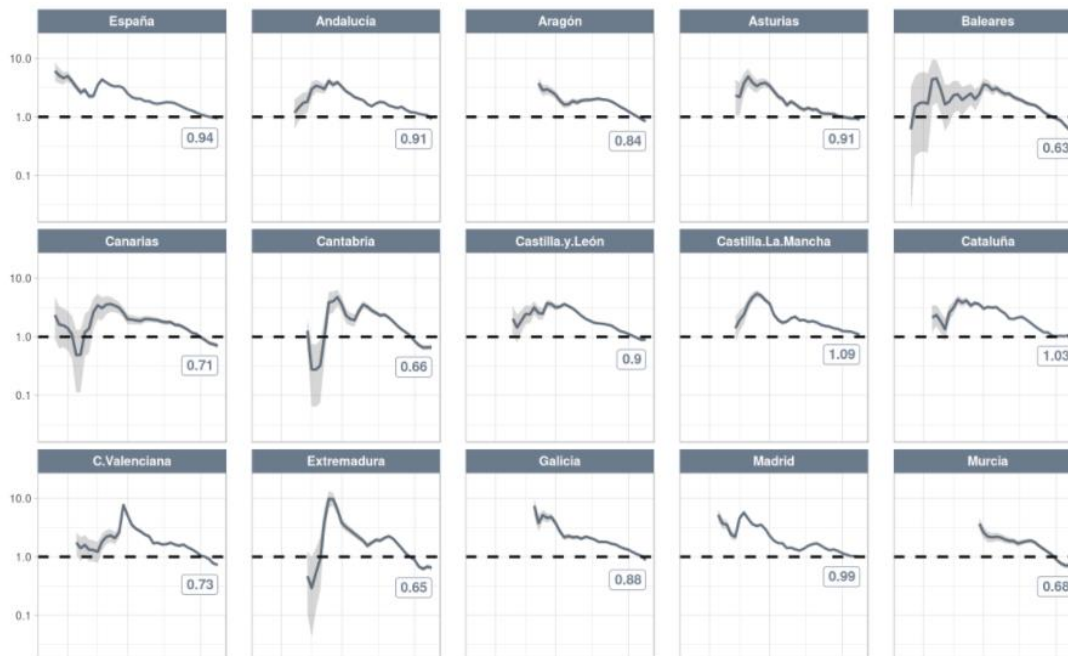


Métricas para cuantificar una difusión

- Obviamente, las tasas de infectados, nuevos infectados, recuperados, etc.
- Pero, uno de los más indicativos, y más de actualidad, es el número reproductivo básico R_0 , que cuantifica la cantidad de nuevo infectados por cada infectado previo.
- Es “sencillo” de medir (ver biblioteca “ R_0 ” de R y los artículos asociados,... igual no lo es tanto).
- Pero es intuitivo: día a día tenemos contactos con varias personas. Si varias de ellas están infectadas, será más probable que nos infecten por lo que, si eso sucede mucho, cada una puede generar más infectados y esto es, precisamente, lo que en media mide R_0 .

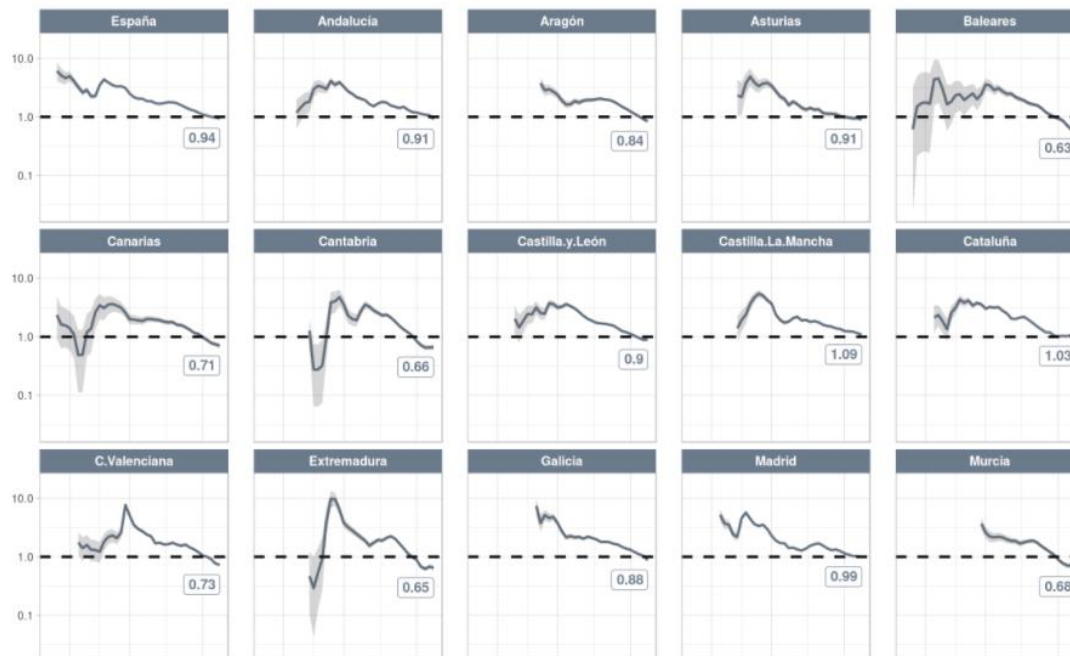
$$R_0 = \beta \langle k \rangle$$

Métricas para cuantificar una difusión



- Sin embargo, a pesar de que esta fórmula es un número que no depende del tiempo, R_0 es una cifra dinámica y varía con el tiempo.
- ¿Por qué sucede esto?

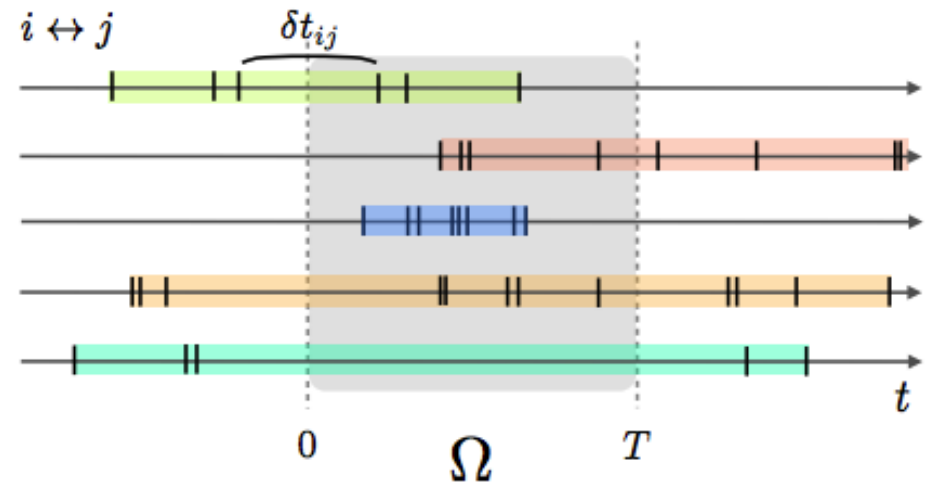
Métricas para cuantificar una difusión



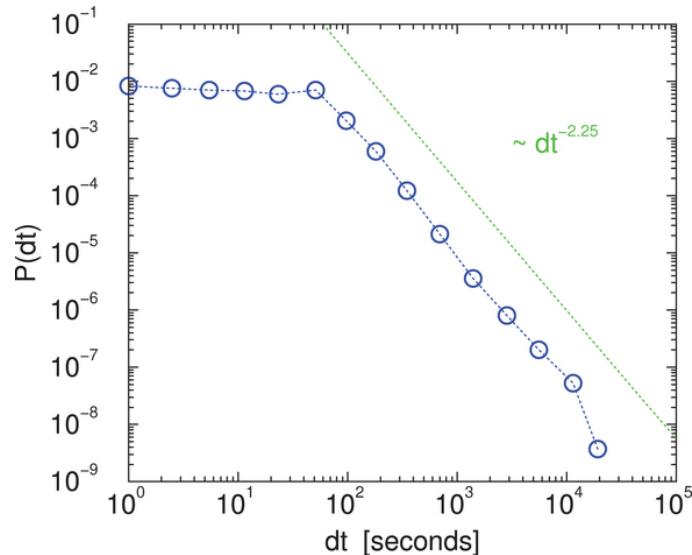
- Sin embargo, a pesar de que esta fórmula es un número que no depende del tiempo, R_0 es una cifra dinámica y varía con el tiempo.
- ¿Por qué sucede esto?
 - Varía el número de gente susceptible
 - Disminuye el número de contactos con personas

Cómo interactuamos modifica la difusión de información

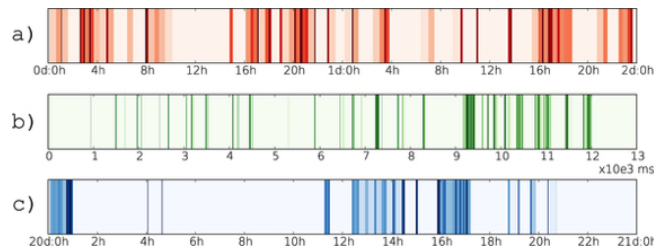
- Para modelizar la difusión de información, siempre consideramos que puede propagarse cuando hay una interacción.
- Decimos que un enlace está vivo en tiempo t si hay al menos una interacción antes de t y al menos otra después.
- Así que hay una manera natural de definir enlaces vivos, enlaces muertos, nuevos, etc.
- ¿Cómo es la distribución del tiempo entre interacciones?



Cómo interactuamos modifica la difusión de información

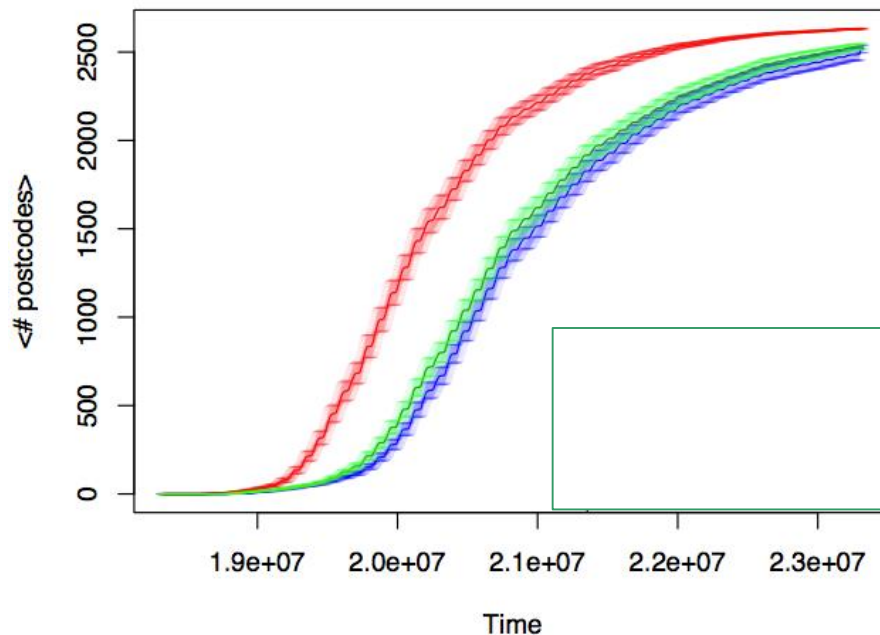


- Típicamente, la distribución entre eventos (interacciones, tweets, etc.) es una Power_law.
- Esto significa que encontramos muchas veces que hay interacciones muy seguidas pero con períodos de inactividad muy largos.
- ¿Cómo afecta esto a la difusión de información?



Cómo interactuamos modifica la difusión de información

- ¿Cómo podemos cuantificar que la difusión de información es debida a la estructura temporal de las interacciones? ¿Si aleatorizamos todos los timestamps de las interacciones, se acelera o se ralentiza la difusión de información?



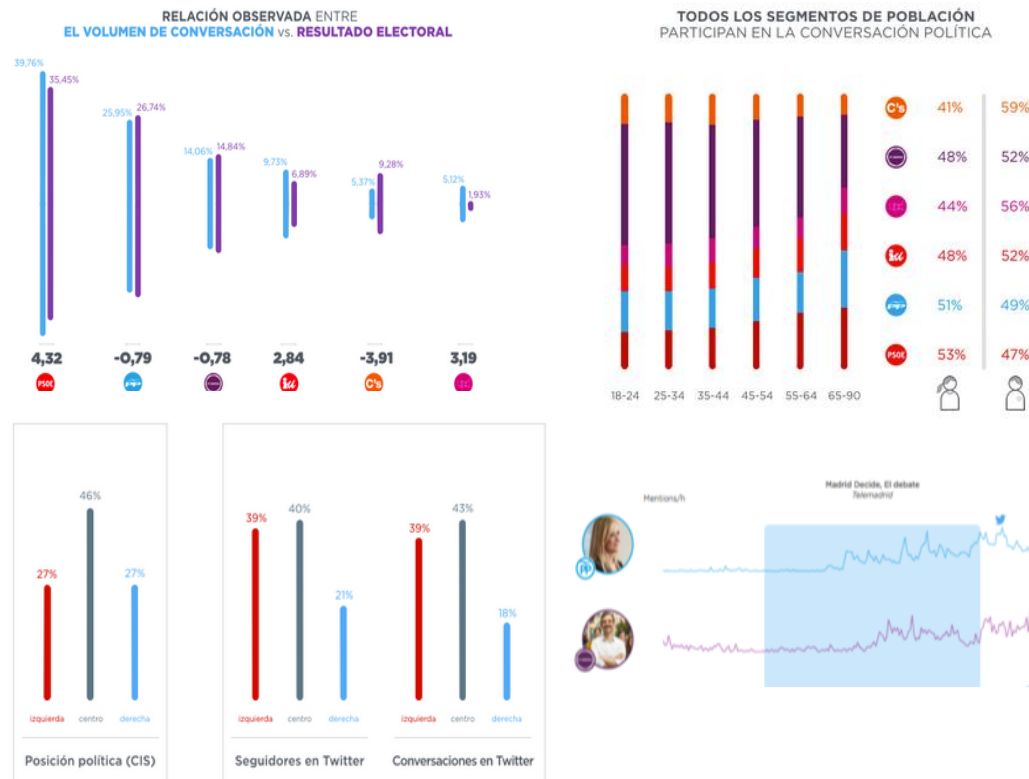
- La media de tiempo entre interacciones es la misma en los dos casos...
- Vemos que la forma en la que interactuamos, ralentiza la difusión de información.
- Podemos probar distintos modelos nulos para aislar el efecto de otras características de la difusión de información como la geografía o la estructura de la red social.

2 | Aplicaciones en redes sociales

Redes sociales en política

Twitter, plataforma clave para movilizar electorado y captar indecisos

Miércoles, Mayo 13, 2015 | Por Alfonso Calatrava (@alcalatrava), Director de Investigación, Twitter



Influencia social acciones políticas



- 611K usuarios recibieron un mensaje en su newsfeed que les animaba a votar.
- 60M de usuarios recibieron el mismo mensaje pero se mostraban 6 amigos aleatorios que habían dado al botón de votar.
- 611K usuarios no recibieron ningún mensaje.

Influencia social acciones políticas



- Cruzaron los datos de cada grupo con datos público de voto de 6.3 millones de votantes.
- Los que vieron el primer mensaje o no vieron ninguno, tuvieron la misma probabilidad de votar.
- Los usuarios que vieron el mensaje con las acciones de sus amigos, aumentaron su probabilidad de voto un 2%. Puede parecer poco, pero esto hizo que **310K personas** votaran en las elecciones (entre directos e indirectos).

Influencia de contenidos sobre emociones



The Guardian

sport football opinion culture business lifestyle fashion environment tech travel [all sections](#)

Facebook apologises for psychological experiments on users

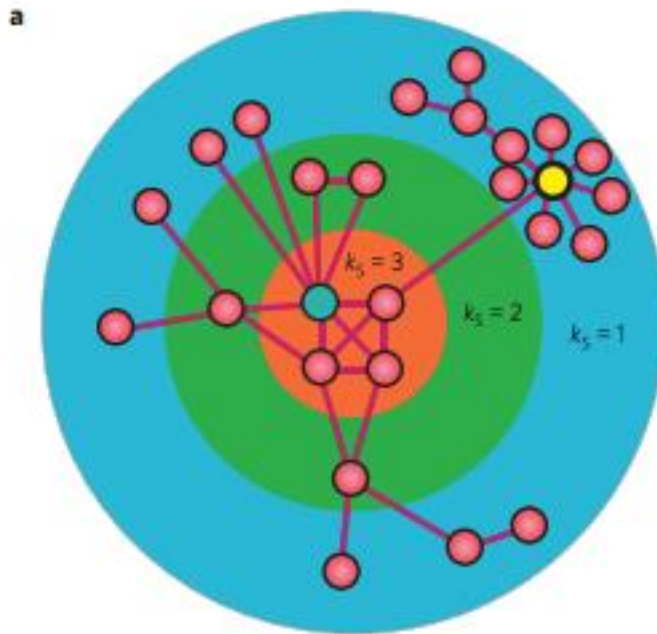
The second most powerful executive at the company, Sheryl Sandberg, says experiments were 'poorly communicated'

¿A quién tengo que mandarle una campaña para que se difunda más?

- En general, las campañas suelen tener objetivos tanto de interacción de los usuarios del target como de visibilidad, es necesario llegar al mayor número de gente posible.
- Sin embargo, el coste de atacar a mucha población suele ser alto (call centers, campañas en redes sociales) aunque no siempre económico (el spam es gratis, pero tiene un daño reputacional).
- ¿Cómo podemos seleccionar de todos los nodos posibles a cuáles tenemos que atacar?

¿A quién tengo que mandarle una campaña para que se difunda más?

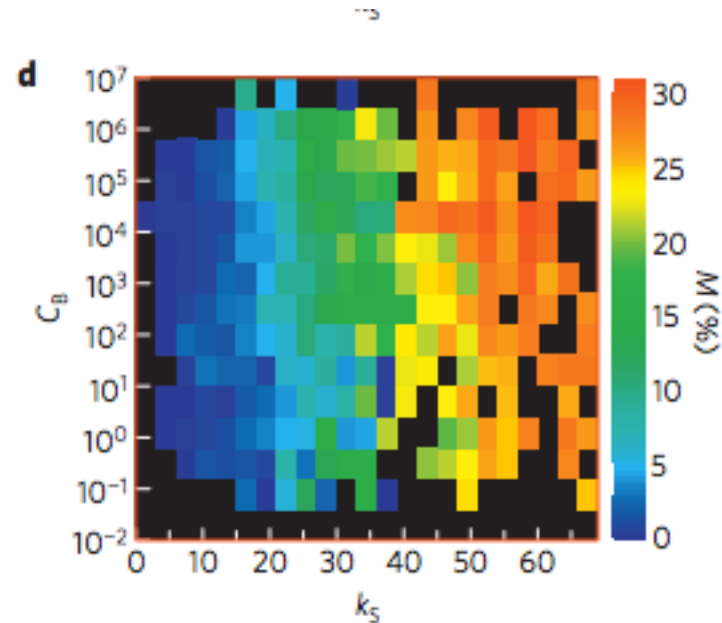
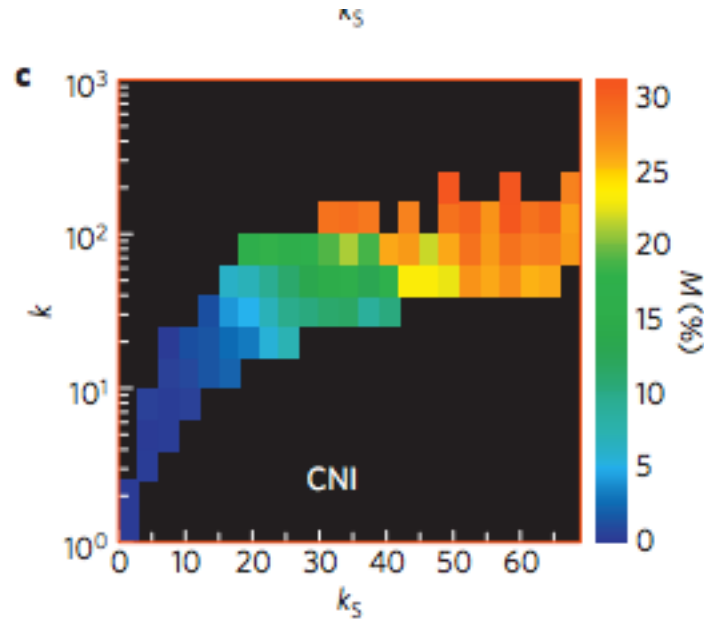
- Uno podría pensar que son los nodos con más grado o los más centrales, por ser hubs o puentes dentro del grafo respectivamente. Sin embargo,...



- No siempre el grado de un nodo determina su nivel de influencia en la propagación de información (el verde y el amarillo tienen el mismo grado)
- K_s representa el nivel de k-core en el que está el nodo (todos los nodos en el nivel S tienen al menos grado S).
- Hay nodos con un gran grado pero que están en la periferia del grafo y no son buenos para difundir información, a pesar de tener muchos contactos.

¿A quién tengo que mandarle una campaña para que se difunda más?

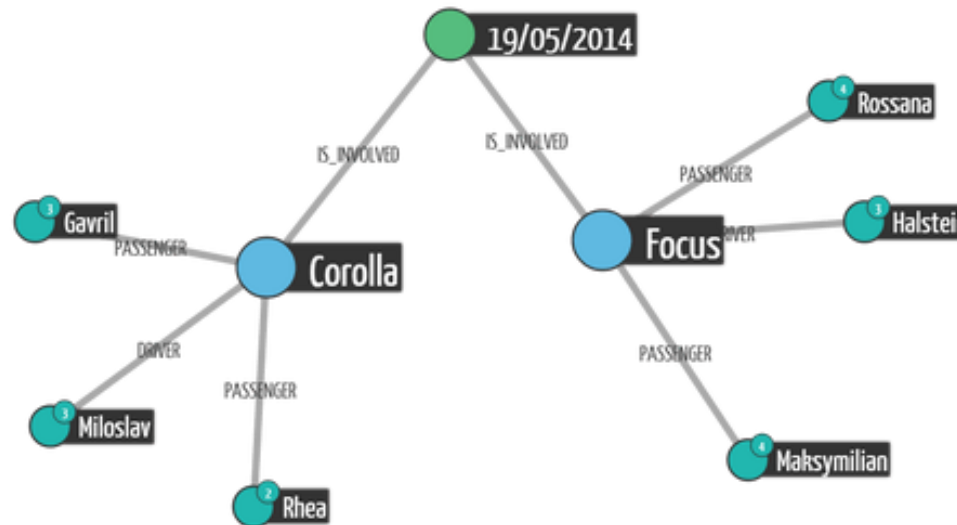
- ¿Cuánta diferencia hay?



Redes sociales en detección de fraude

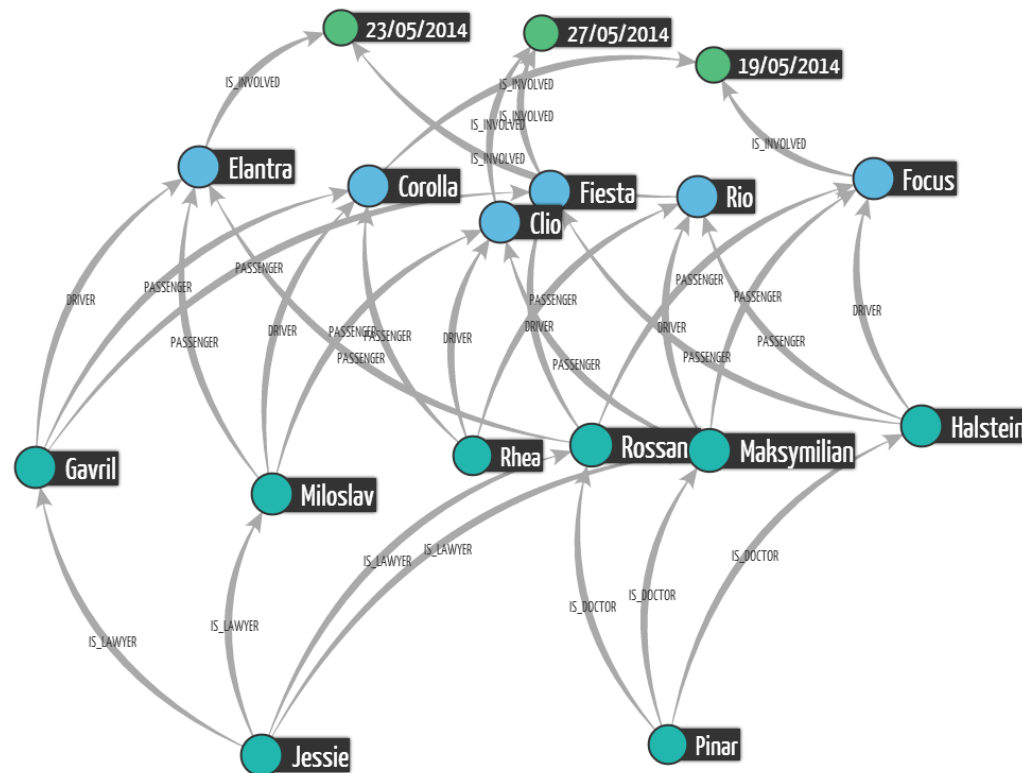
Podemos utilizar técnicas de análisis de redes sociales para detectar fraude en aseguradoras.

Imaginamos un escenario donde ha habido un accidente. Esto mismo lo podemos representar en una red múltiple donde los nodos representan diferentes tipos de entidades y los enlaces también.



Redes sociales en detección de fraude

Consultamos en una base de datos de grafo (hay implementaciones que hacen esto rápido, como Neo4J) y vemos que aparecen dos accidentes más y que hay una sospechosa relación entre los involucrados en los accidentes...



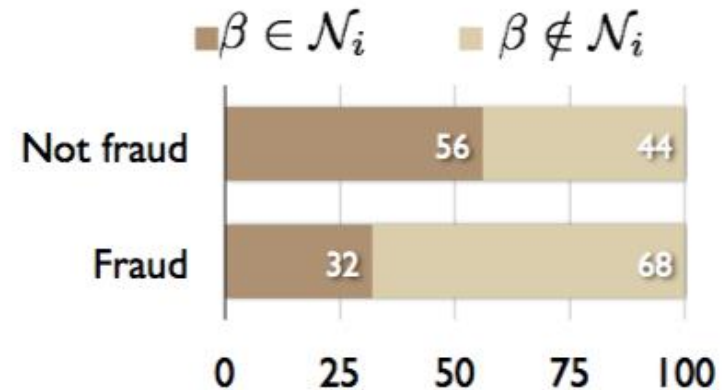
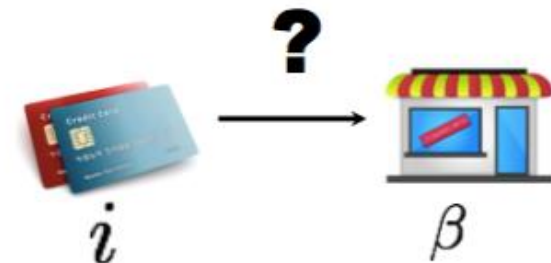
Redes sociales en detección de fraude

- Utilizar nuestros movimientos para detectar robos/fraude con tarjeta de crédito



$$\beta \in \mathcal{N}_i = \{\alpha_1, \dots, \alpha_{n_1}\}$$

$$\beta \notin \mathcal{N}_i = \{\alpha_1, \dots, \alpha_{n_1}\}$$



Aproximando el nivel educacional con datos de redes sociales

- ¿Está relacionada la complejidad de los tweets con el nivel educacional de las regiones? Solucionar esto es un problema relacionado con la economía de una región, ya que aquellas regiones con mayor nivel educacional tienden a tener mejores niveles económicos.

The Readability of Tweets and their Geographic Correlation with Education

James R. A. Davenport
University of Washington
Seattle, WA
jrad@astro.washington.edu

Robert DeLine
Microsoft Research
Redmond, WA
rdeline@microsoft.com

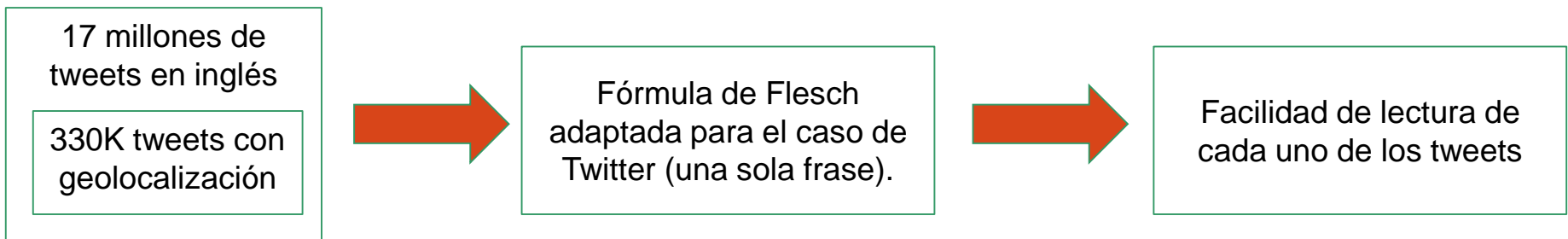
ABSTRACT

Twitter has rapidly emerged as one of the largest worldwide venues for written communication. Thanks to the ease with which vast quantities of tweets can be mined, Twitter has also become a source for studying modern linguistic style. The readability of text has long provided a simple method to characterize the complexity of language and ease that documents may be understood by readers. In this note we use a modified version of the Flesch Reading Ease formula, applied to a corpus of 17.4 million tweets. We find tweets have characteristically more difficult readability scores

new mediums such as Twitter.

The “readability” of text has long been used to predict the difficulty people will have in understanding written content. A variety of approaches have historically been used to calculate readability scores. These range from counting the density of previously identified “difficult” words, to various schemes of counting the number of syllables per word and words per sentence. Many public agencies require that laws and official documents conform to thresholds of readability or “Reading Ease” scores. This is intended to promote and

Aproximando el nivel educacional con datos de redes sociales



- Comparación con SMS y Chat.
- La media del RE para Twitter fue de 50 mientras que para los datos de chats es de 54 (muy similar a los datos de Twitter) y bastante menor que los datos de SMS, con una media de 88 (es más fácil leer SMS que tweets).

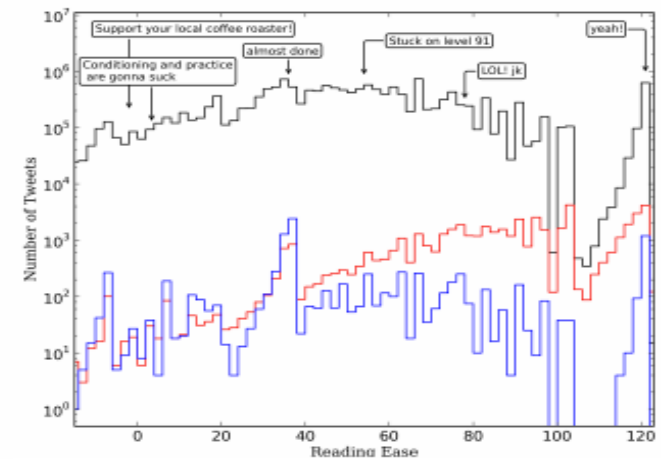


Figure 1. Histograms of RE for our three samples, 17.4 million tweets (black), the NUS SMS corpus (red), and the NPS chat corpus (blue). Higher RE values are easier to read. Example tweets are shown, with arrows indicating their RE score.

Aproximando el nivel educacional con datos de redes sociales

- ¿Existe correlación entre los niveles educacionales oficiales y la complejidad de los tweets de una región?

330K tweets con geolocalización

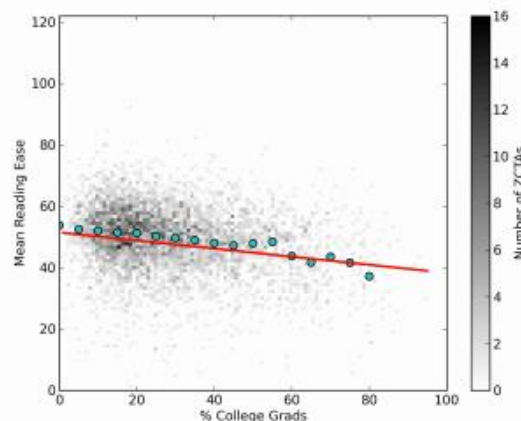


Fórmula de Flesch adaptada para el caso de Twitter (una sola frase).



Facilidad de lectura de cada uno de los tweets

- Correlación del valor medio del RE frente al porcentaje de graduados en la región.



$$\beta = -0.15$$

- Es decir, a menor complejidad en la escritura, menor nivel educacional.

Aproximando el nivel educacional con datos de redes sociales

- Pensando en lo bien o mal que escriben los usuarios de redes sociales, se nos pueden ocurrir utilizar otro tipo de variables...

Tweet	Detected Misspellings
Alguien se viene con migo aver la vida de PI??	<ul style="list-style-type: none">- "Con migo" instead of "Conmigo" (<i>with me</i> in Spanish).- "aver" instead of "a ver" (<i>aver</i> is not a Spanish word)
La quiero mucho y la hecho de menos	<ul style="list-style-type: none">- "Hecho de menos" instead of "echo de menos" (<i>"I miss her"</i> in Spanish).

All the 618 expressions such as "Con migo", "Aver" or "Hecho de menos" have been searched literally within the text of the whole dataset of tweets.

Aproximando el nivel educacional con datos de redes sociales

- La complejidad léxica es un proxy del nivel educacional.
- Escribir mal... también.

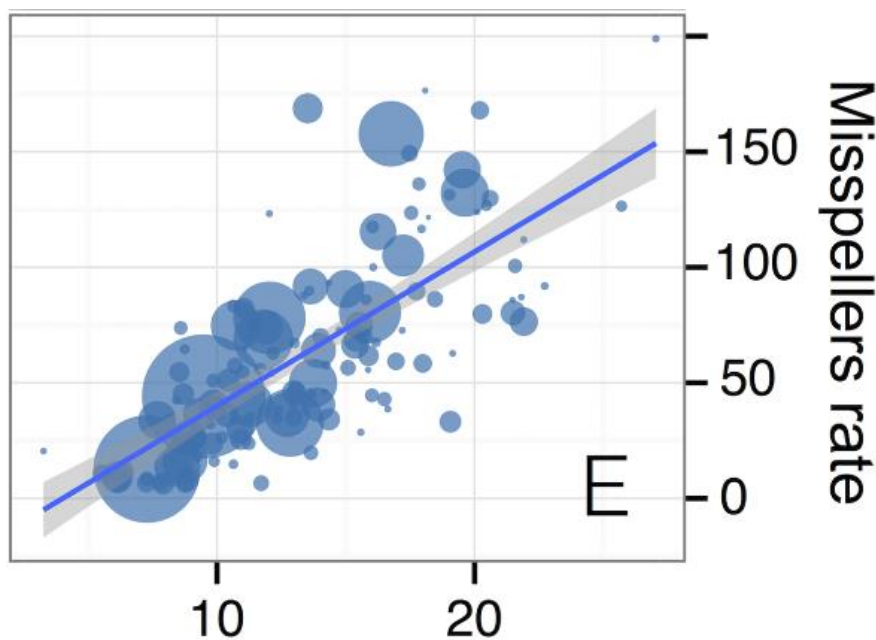
Serious misspellings:

- ❖ haber que tal
- ❖ Aprovadas
- ❖ Cunplimos
- ❖ ...

- No tenemos en cuenta seseos (o similares), abreviaturas... solo buscamos las mayores burradas.
- Contamos el porcentaje de usuarios de Twitter que cometen faltas de este estilo en las regiones donde queremos modelizar fenómenos económicos.

Aproximando el nivel educacional con datos de redes sociales

- La proporción de usuarios en una región que cometen este tipo de faltas está muy correlacionada con el paro.



$$\beta = 0.65$$

Midiendo el mercado laboral con datos de social media

- Aparte del porcentaje de usuarios que cometen faltas de ortografía, ¿qué otras variables podemos extraer de Twitter que nos permiten modelizar el paro?

Social Media Fingerprints of Unemployment

Alejandro Llorente, Manuel Garcia-Herranz, Manuel Cebrian, Esteban Moro 

Published: May 28, 2015 • <http://dx.doi.org/10.1371/journal.pone.0128692>

Article	Authors	Metrics	Comments	Related Content
				

Abstract

Introduction

Results

Discussion

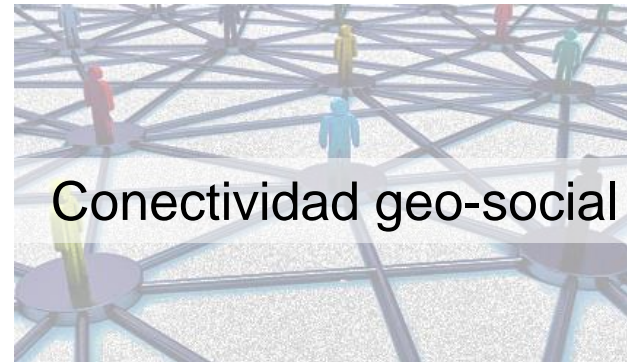
Supporting Information

Acknowledgments

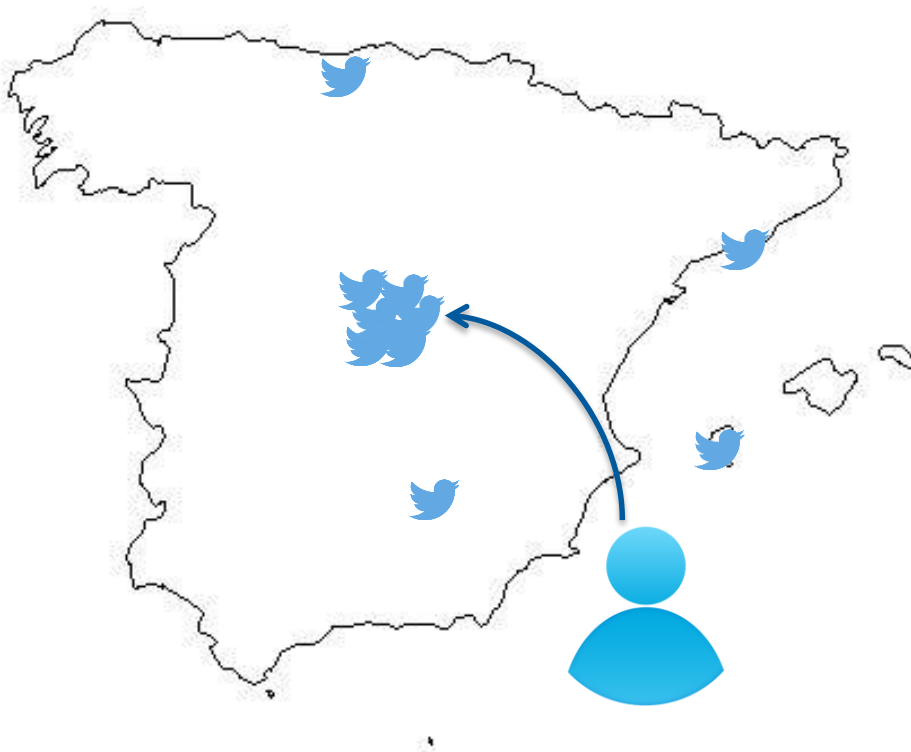
Abstract

Recent widespread adoption of electronic and pervasive technologies has enabled the study of human behavior at an unprecedented level, uncovering universal patterns underlying human activity, mobility, and interpersonal communication. In the present work, we investigate whether deviations from these universal patterns may reveal information about the socio-economical

Midiendo el mercado laboral con datos de social media

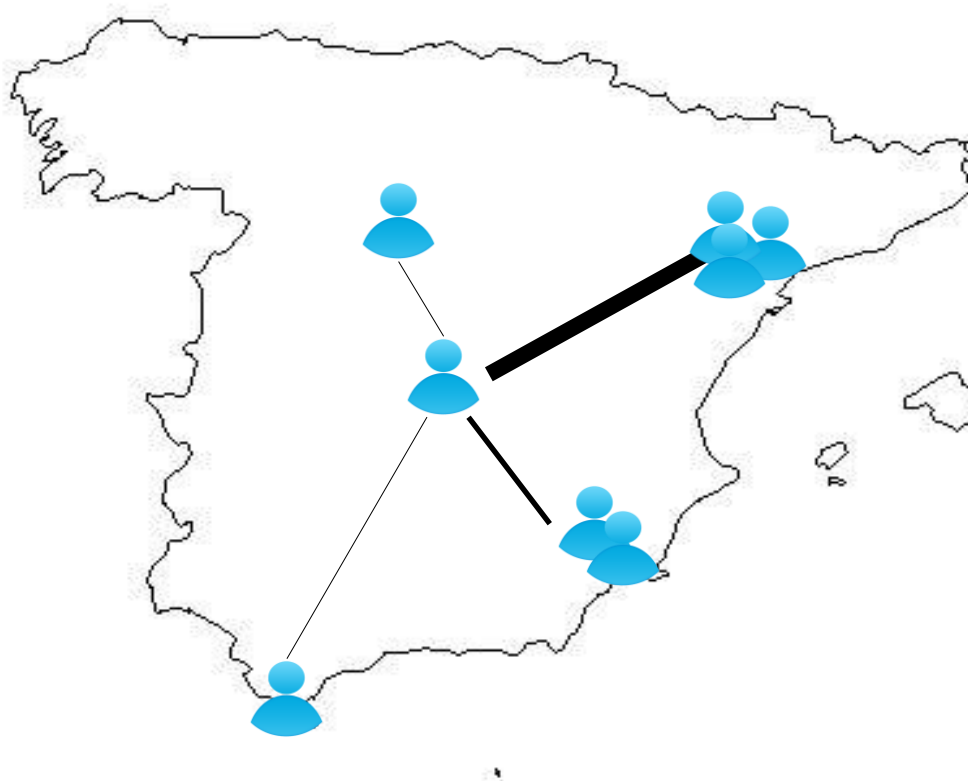


Midiendo el mercado laboral con datos de social media



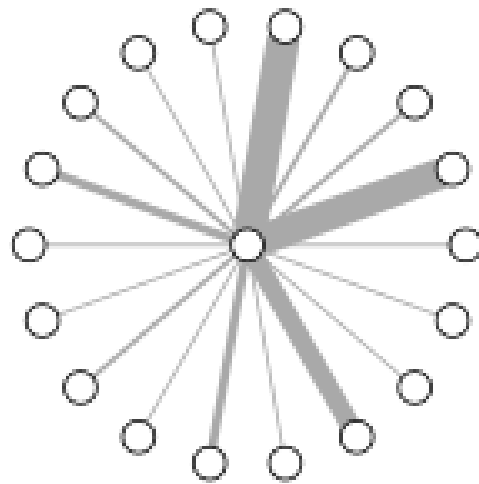
- Correlación entre número de usuarios y población = 0.96
- La penetración es el número de usuarios sobre el total de la población.
- Trabajos previos vieron que en países, a más penetración mayor desarrollo económico.

Midiendo el mercado laboral con datos de social media

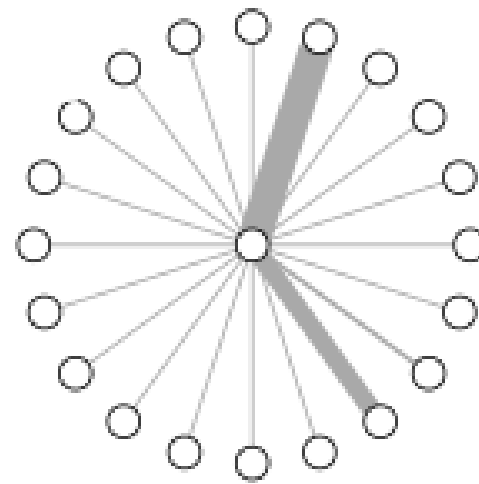


- Para cada usuario, medimos la diversidad de sitios con los que interactúa mediante menciones.
- La conectividad geo-social de una comunidad, es la media de las conectividades de los usuarios que viven en ella.

Midiendo el mercado laboral con datos de social media

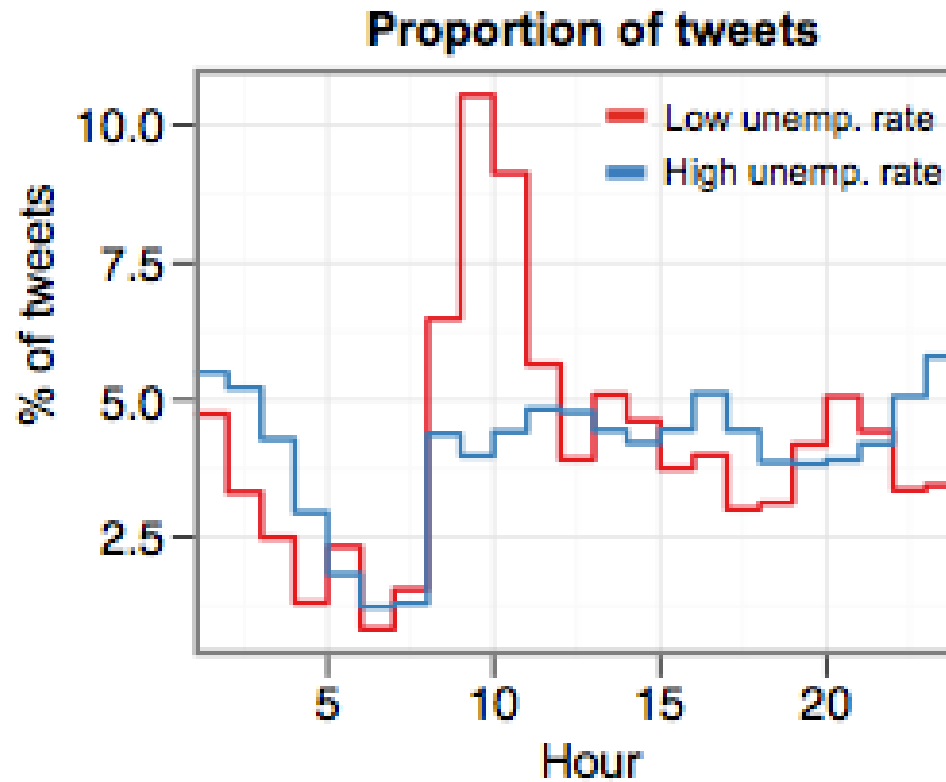


Entropy: **0.72**
Unemployment rate: **11%**



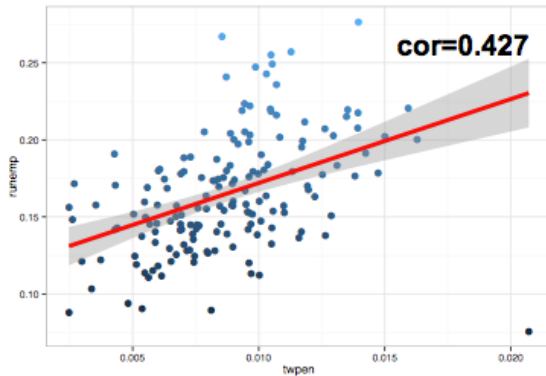
Entropy: **0.42**
Unemployment rate: **23%**

Midiendo el mercado laboral con datos de social media

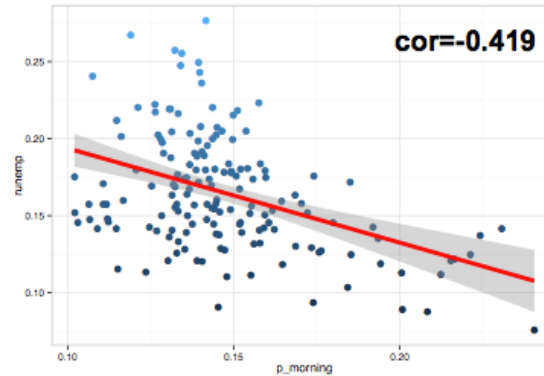


Midiendo el mercado laboral con datos de social media

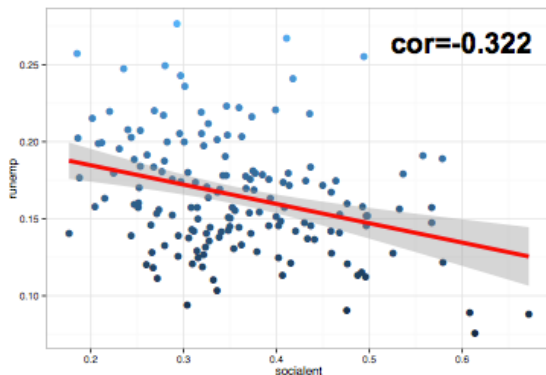
Unemp. vs Penetration Rate



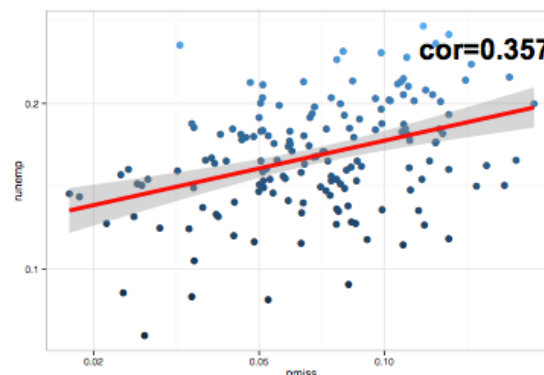
Unemp. vs Morning Activity



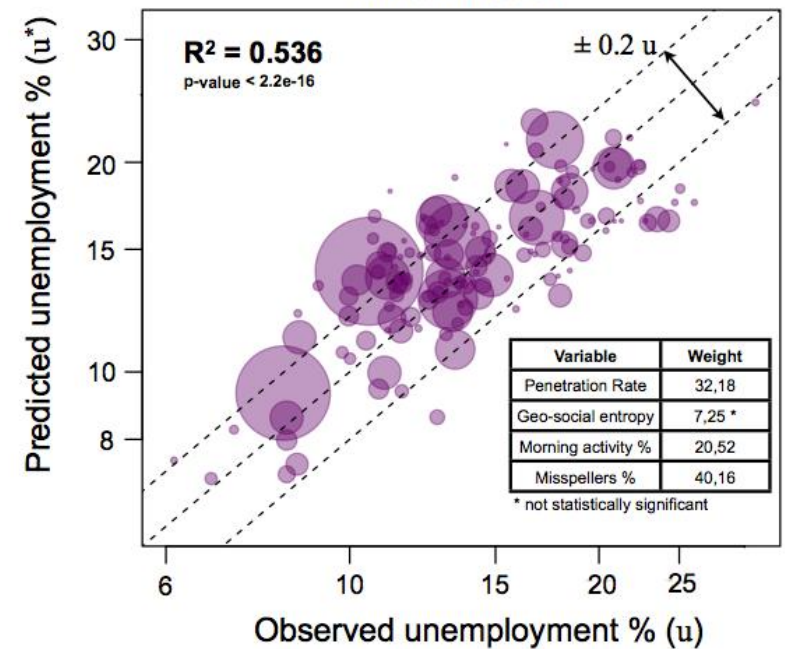
Unemp. vs Social Entropy



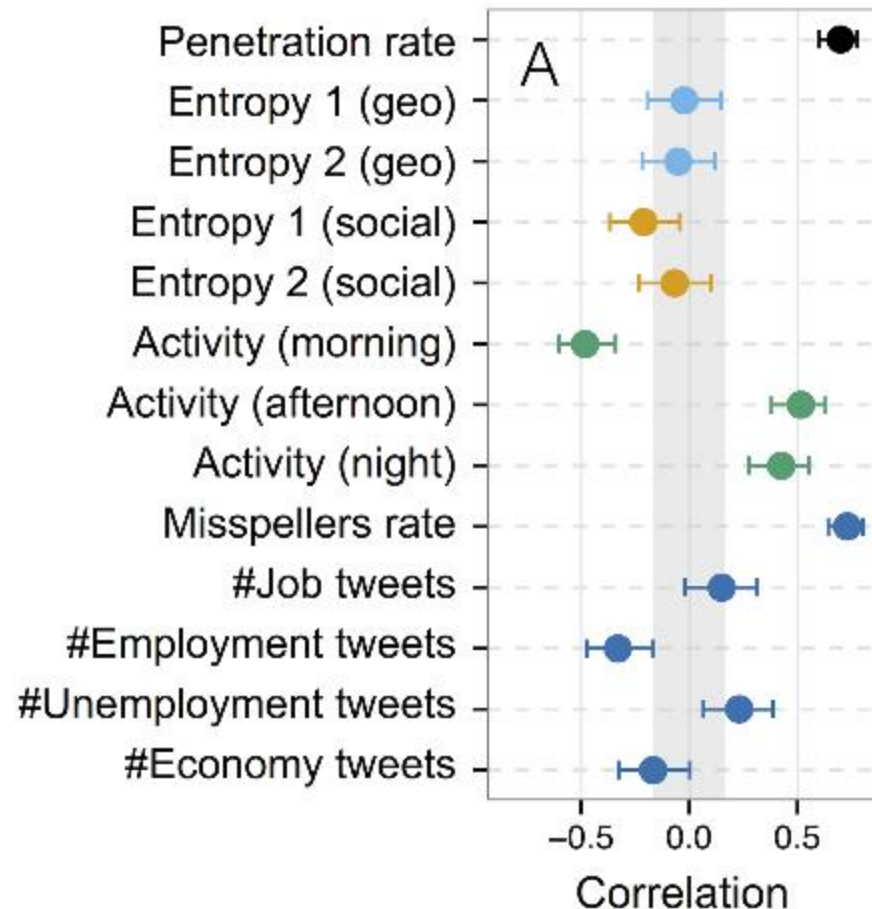
Unemp. vs Misspellers Rate



Linear Model



Midiendo el mercado laboral con datos de social media



Utilizando logs de búsquedas para cuantificar comportamientos financieros

- ¿Tiene alguna información las diferencias de volúmenes de búsquedas en Google con los mercados financieros?

Quantifying Trading Behavior in Financial Markets Using Google Trends

Tobias Preis , Helen Susannah Moat & H. Eugene Stanley

Scientific Reports **3**, Article number: 1684 (2013)

doi:10.1038/srep01684

[Download Citation](#)

Applied physics Computational science

Information theory and computation

Statistical physics, thermodynamics and

Received: 25 February 2013

Accepted: 03 April 2013

Published online: 25 April 2013

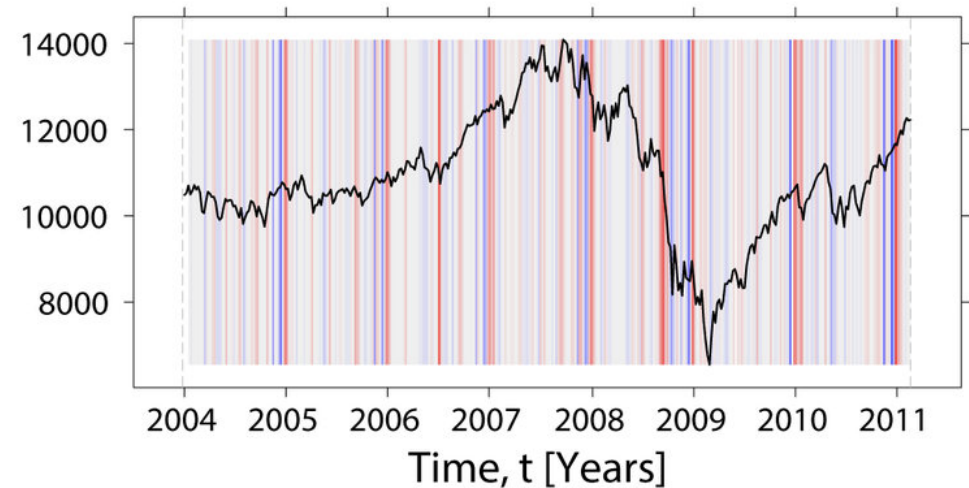
Utilizando logs de búsquedas para cuantificar comportamientos financieros

- Hipótesis: la variación del volumen de búsquedas de términos financieros está correlacionada con los futuros eventos que se observan en los mercados financieros. Si esto es cierto, podremos crear estrategias ganadoras en bolsa.

Serie temporal del valor del DJIA al inicio de la semana



Apariciones del términos “deuda” en las búsquedas junto a la evolución del DJIA



Utilizando logs de búsquedas para cuantificar comportamientos financieros

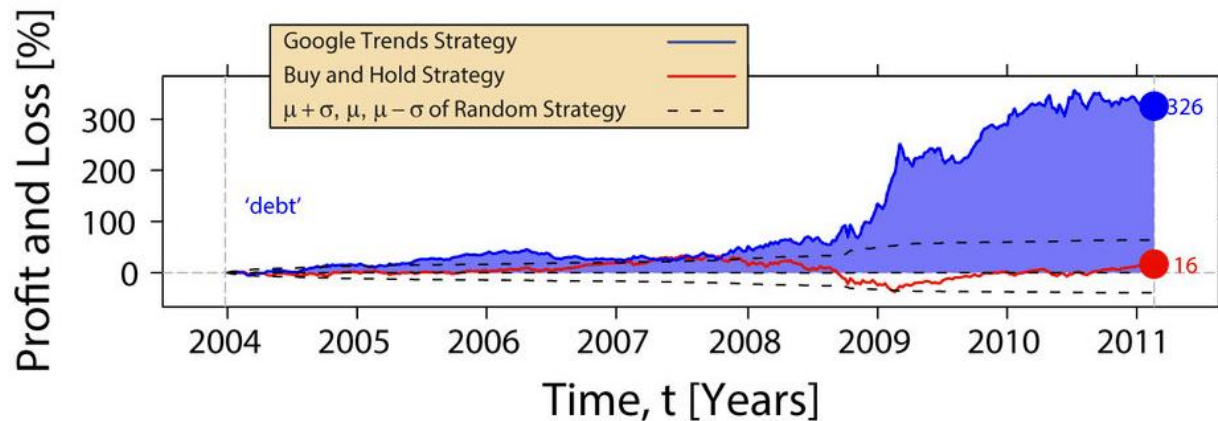
- Comparamos el volumen de búsquedas en la semana $t-1$ con el precio del DJIA en tiempo t .
- Calculamos la variación en el volumen de búsquedas respecto a la media de las últimas semanas:

$$\Delta n(t, \Delta t) = n(t) - \frac{(n(t-1) + \dots + n(t-\Delta t))}{\Delta t}$$

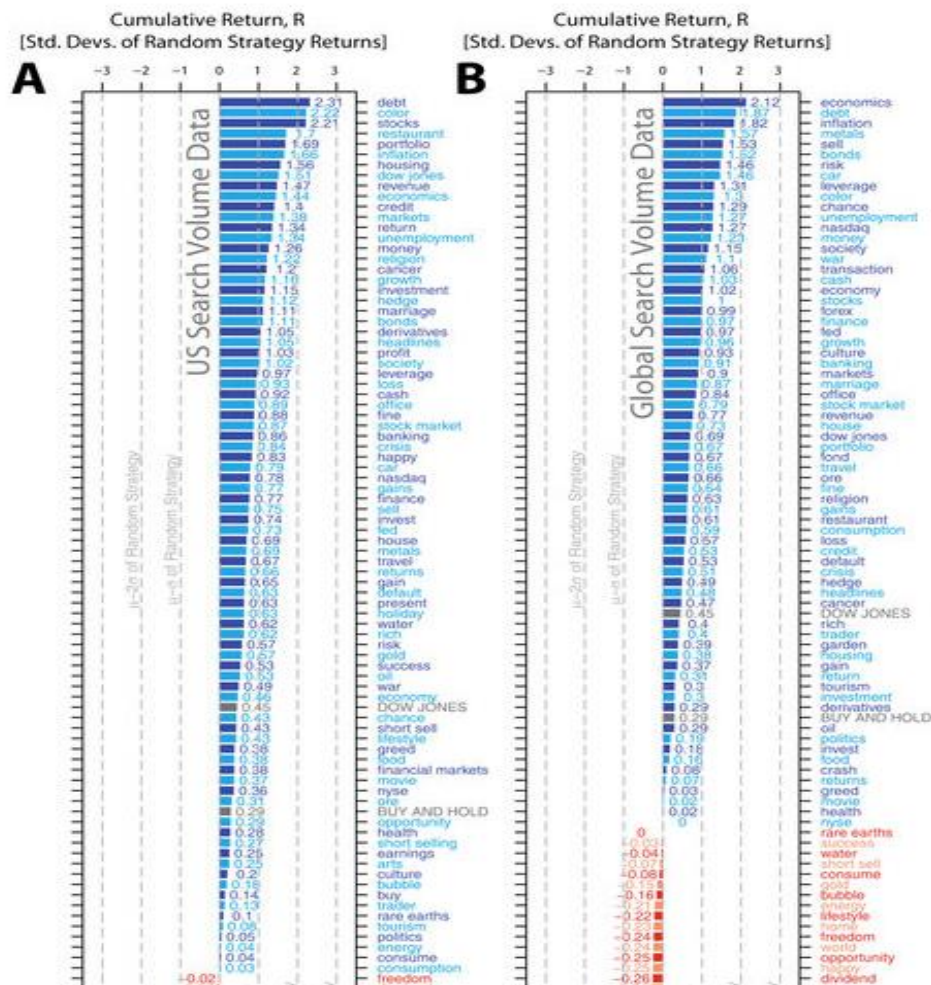
Utilizando logs de búsquedas para cuantificar comportamientos financieros

- Implementamos la siguiente estrategia:
 - Si $\Delta n(t-1, \Delta t) > 0$, entonces vendemos DJIA al precio del día t y compramos en el día $t+1$ (esperamos que el precio baje en las siguientes semanas).
 - Por el contrario, si $\Delta n(t-1, \Delta t) < 0$, compramos DJIA al precio del día t y vendemos en el día $t+1$ (esperamos que el precio suba en las siguientes semanas).

Rendimiento acumulado para la estrategia con la palabra “deuda”



Utilizando logs de búsquedas para cuantificar comportamientos financieros




La ventaja de buscar sobre el futuro

- ¿Reflejan las búsquedas sobre el futuro en Google el estado económico de un país?

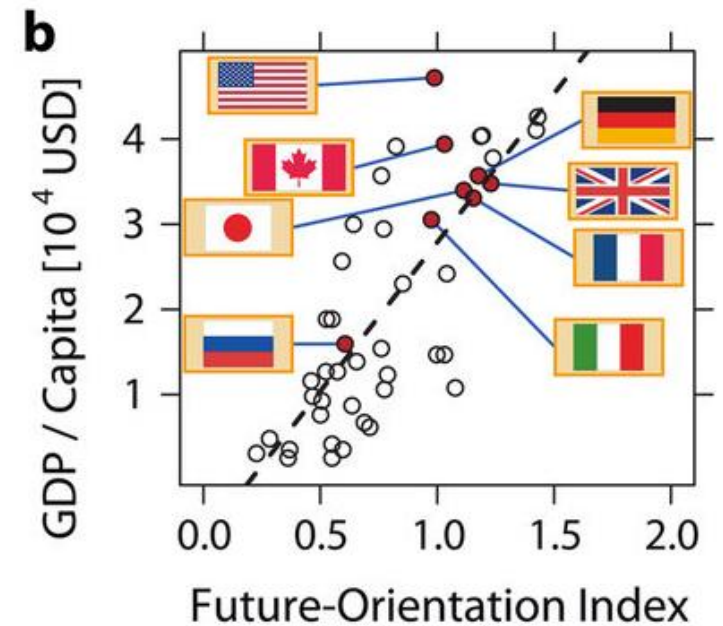
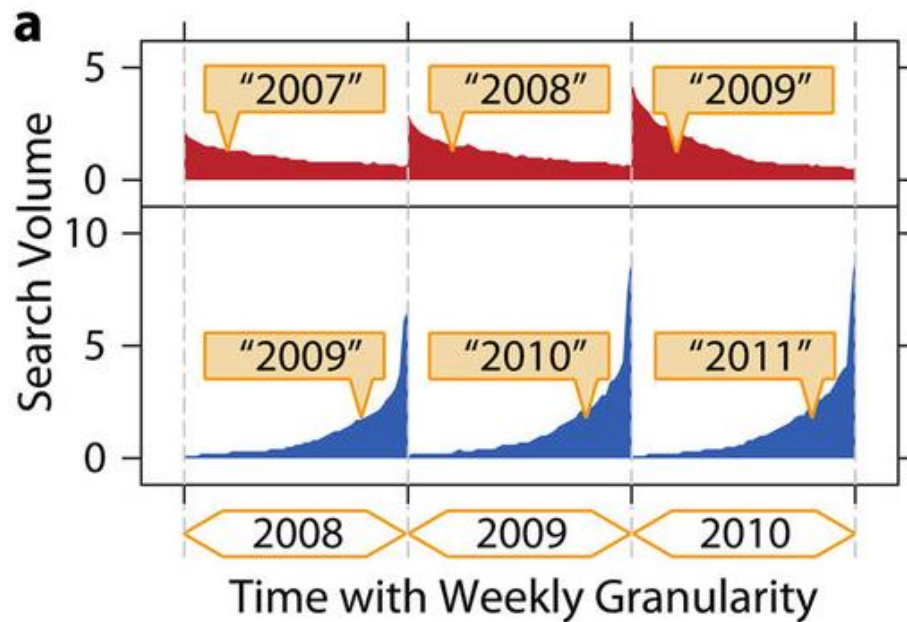
$$\text{Future Orientation Index} = \frac{\text{Número de búsquedas sobre el futuro}}{\text{Número de búsquedas sobre el pasado}}$$

- La manera en la que podemos identificar si una búsqueda es sobre el futuro o el pasado lo podemos hacer simplemente analizando si aparece un año venidero o pasado en la búsqueda y descartando todos los datos que no hacen referencia a un año.

Quantifying the Advantage of Looking Forward

Tobias Preis , Helen Susannah Moat, H. Eugene Stanley & Steven R. Bishop

La ventaja de buscar sobre el futuro



$$R^2 = 0.78$$

Estimando daños de catástrofes naturales con social media

- Anticipar el daño de catástrofes permite a las aseguradoras aprovisionar los fondos necesarios para cubrir todos los gastos derivados de las mismas.
- ¿Permiten los datos de redes sociales anticipar este daño?

Rapid assessment of disaster damage using social media activity

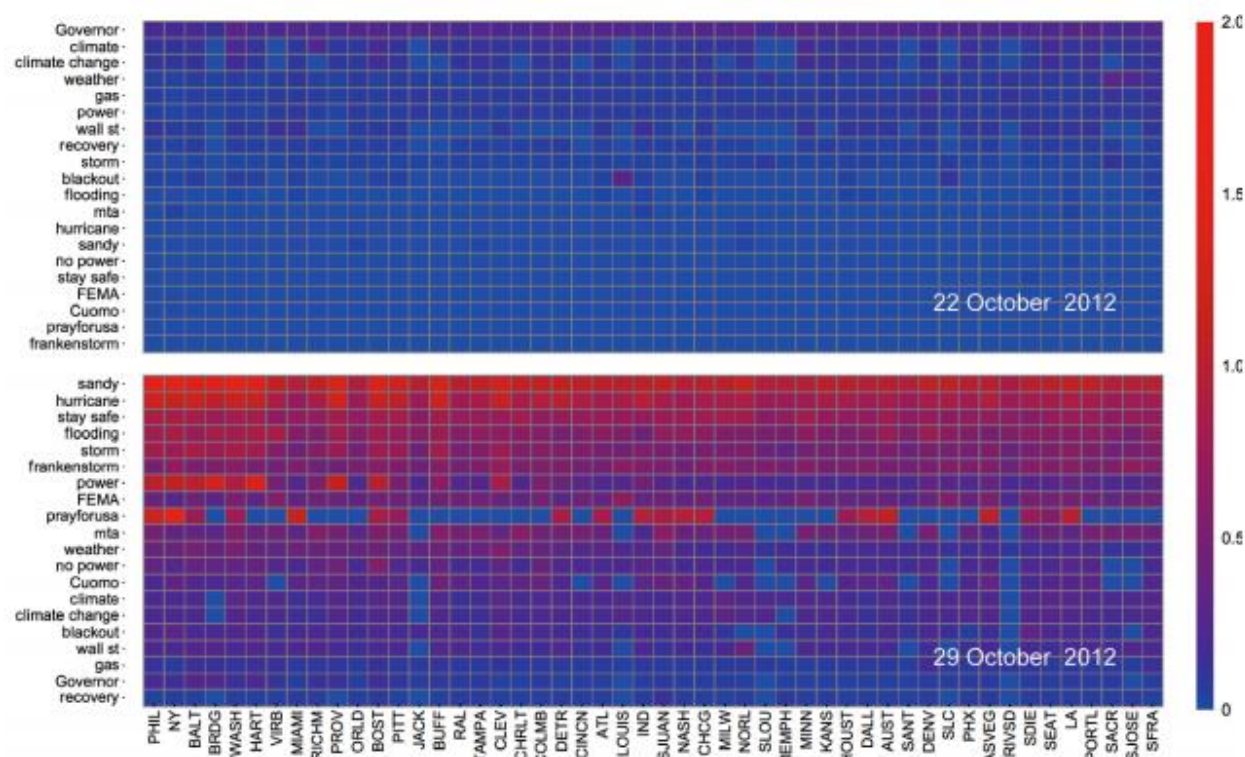
Yury Kryvasheyeu,^{1,2,3} Haohui Chen,^{1,2,3} Nick Obradovich,^{4,5} Esteban Moro,⁶ Pascal Van Hentenryck,¹ James Fowler,^{4,9} Manuel Cebrian^{1,3,10*}

exclusive licensee Ar
the Advancement o
under a Creative Co
NonCommercial Lice
10.1126/scladv.15001

Could social media data aid in disaster response and damage assessment? Countries face both an increasing frequency and an increasing intensity of natural disasters resulting from climate change. During such events, citizens turn to social media platforms for disaster-related communication and information. Social media improves situational awareness, facilitates dissemination of emergency information, enables early warning systems, and helps coordinate relief efforts. In addition, the spatiotemporal distribution of disaster-related messages helps with the real-time monitoring and assessment of the disaster itself. We present a multiscale analysis of Twitter activity before, during, and after Hurricane Sandy. We examine the online response of 50 metropolitan areas of the United States and find a strong relationship between proximity to Sandy's path and hurricane-related social media activity. We show that real and perceived threats, together with physical disaster effects, are directly observable through the intensity and composition of Twitter's message stream. We demonstrate that per-capita Twitter activity strongly correlates with the per-capita economic damage inflicted by the hurricane. We verify our findings for a wide range of disasters and suggest that massive online social networks can be used for rapid assessment of damage caused by a large-scale disaster.

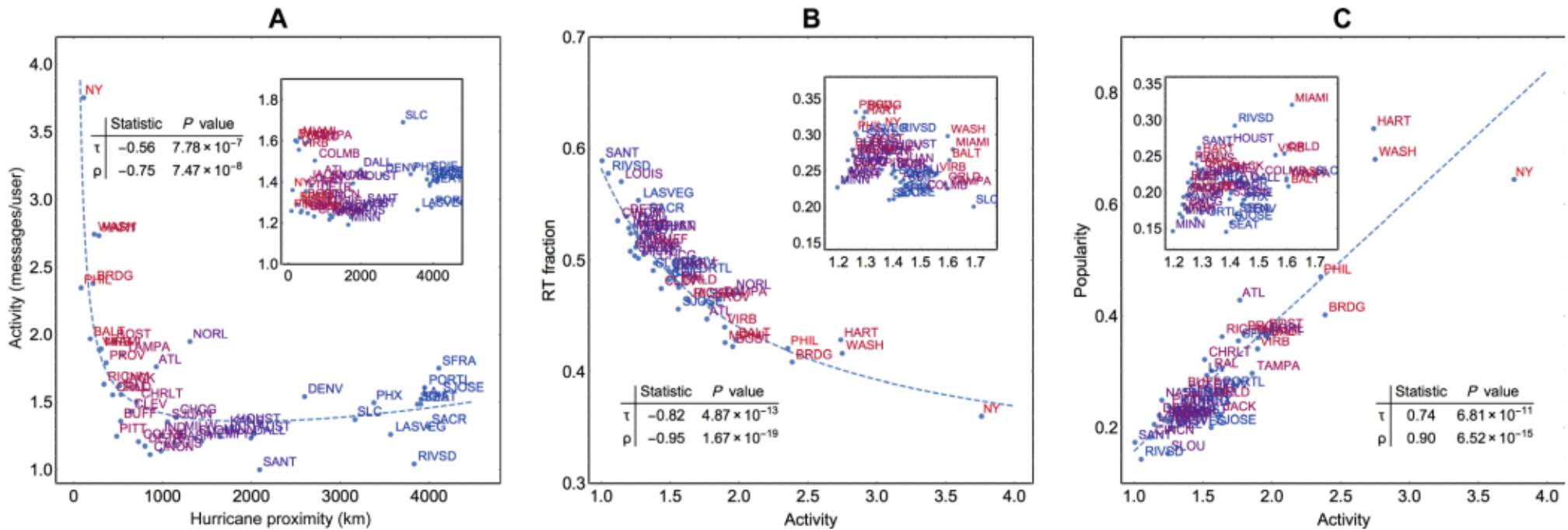
Estimando daños de catástrofes naturales con social media

- Se ve que las ciudades más próximas al huracán tienen una mayor presencia de palabras relacionadas.



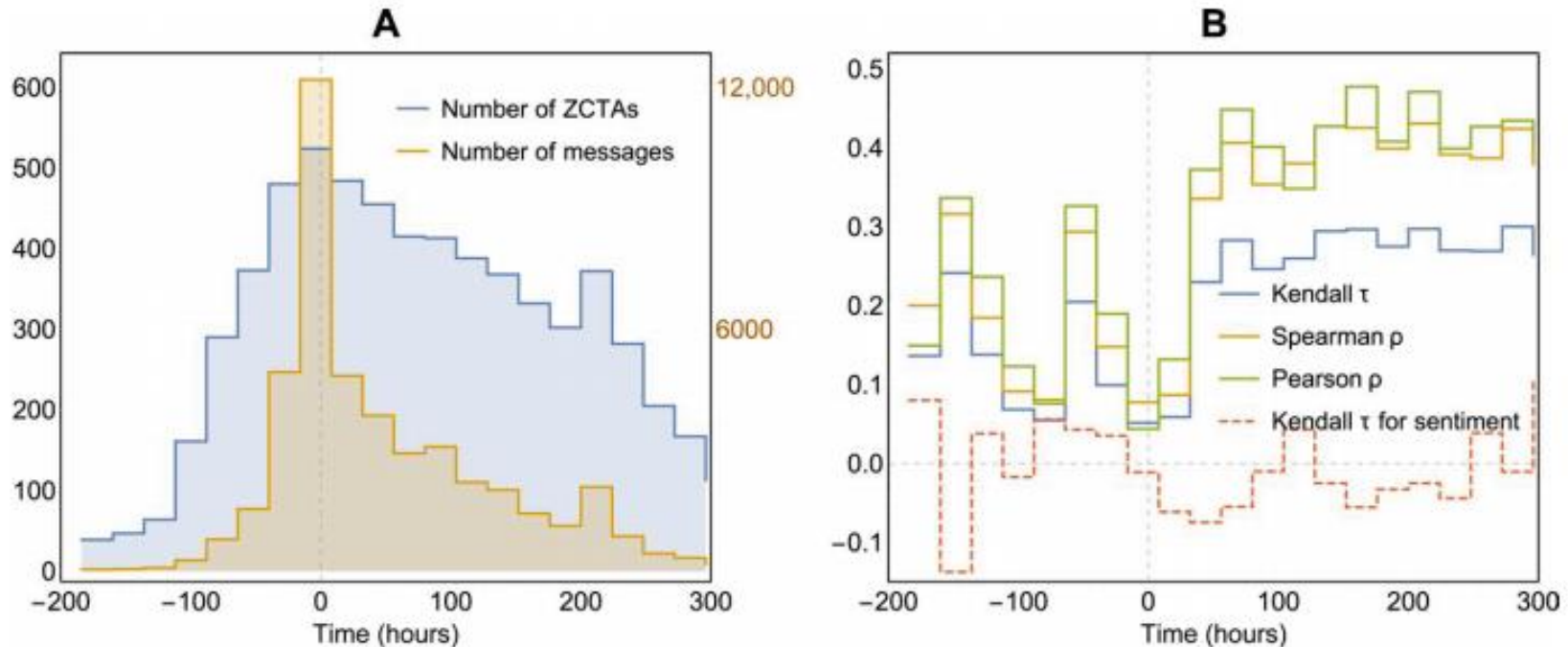
Estimando daños de catástrofes naturales con social media

- En las zonas más cercanas al huracán, existe una mayor actividad y una mayor creación de contenidos originales que en zonas más alejadas y también sus contenidos se difunden más.



Estimando daños de catástrofes naturales con social media

- Finalmente, la actividad horaria en redes sociales relacionadas con la catástrofe natural, correla con el número de solicitudes recibidas por las aseguradoras, especialmente la actividad en social media de las horas después al momento de máximo auge el huracán.



3 | Práctica evaluable



Afi Escuela
de Finanzas

© 2022 Afi Escuela de Finanzas. Todos los derechos reservados.