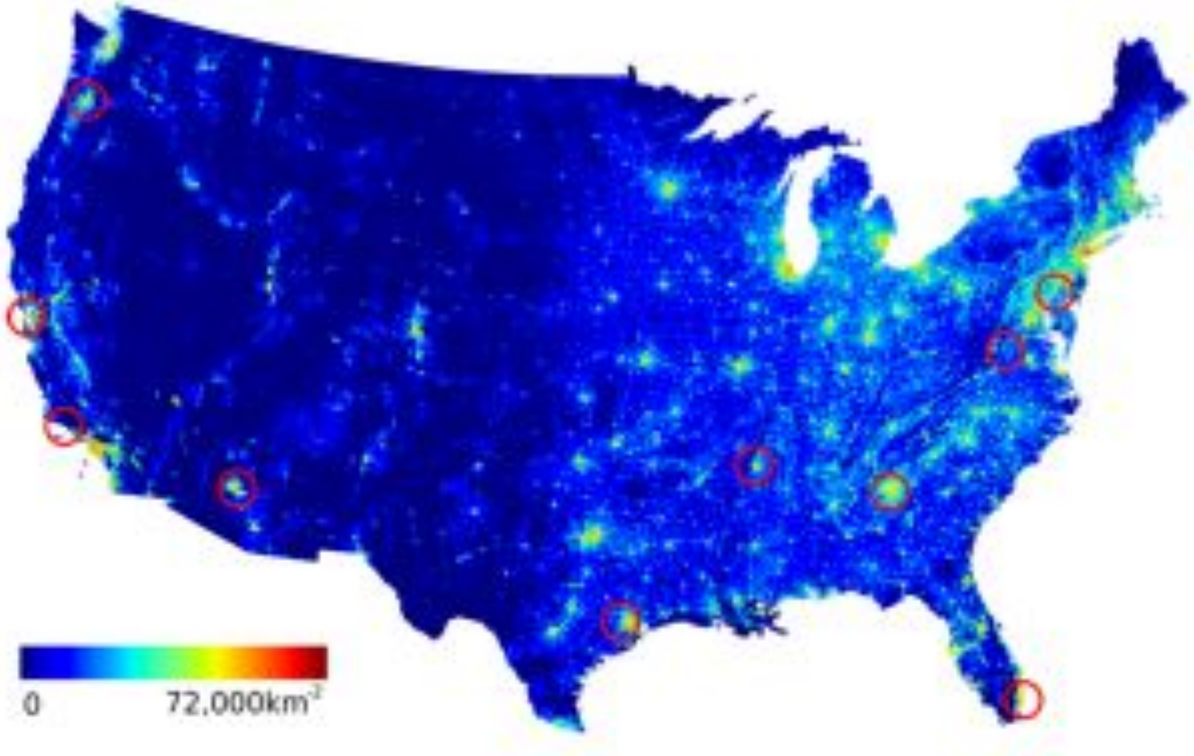
# Following **information in social networks:** how social **structure** and **dynamics** impact message transmission

Esteban Moro Egido
UC3M + IIC

Universidad Carlos III de Madrid

iic instituto de ingeniería del conocimiento

VIII Jornadas MAVIR




Manuel Cebrián
NICTA Australia


Alex "Sandy" Pentland
MIT


Alex Rutherford
United Nations


Przemek Grabowicz
MaxPlanck


Alejandro Llorente
IIC


Manuel Garcia-Herranz
UAM


Iyad Rahwan
Masdar Inst.


Giovanna Miritello
Telefónica

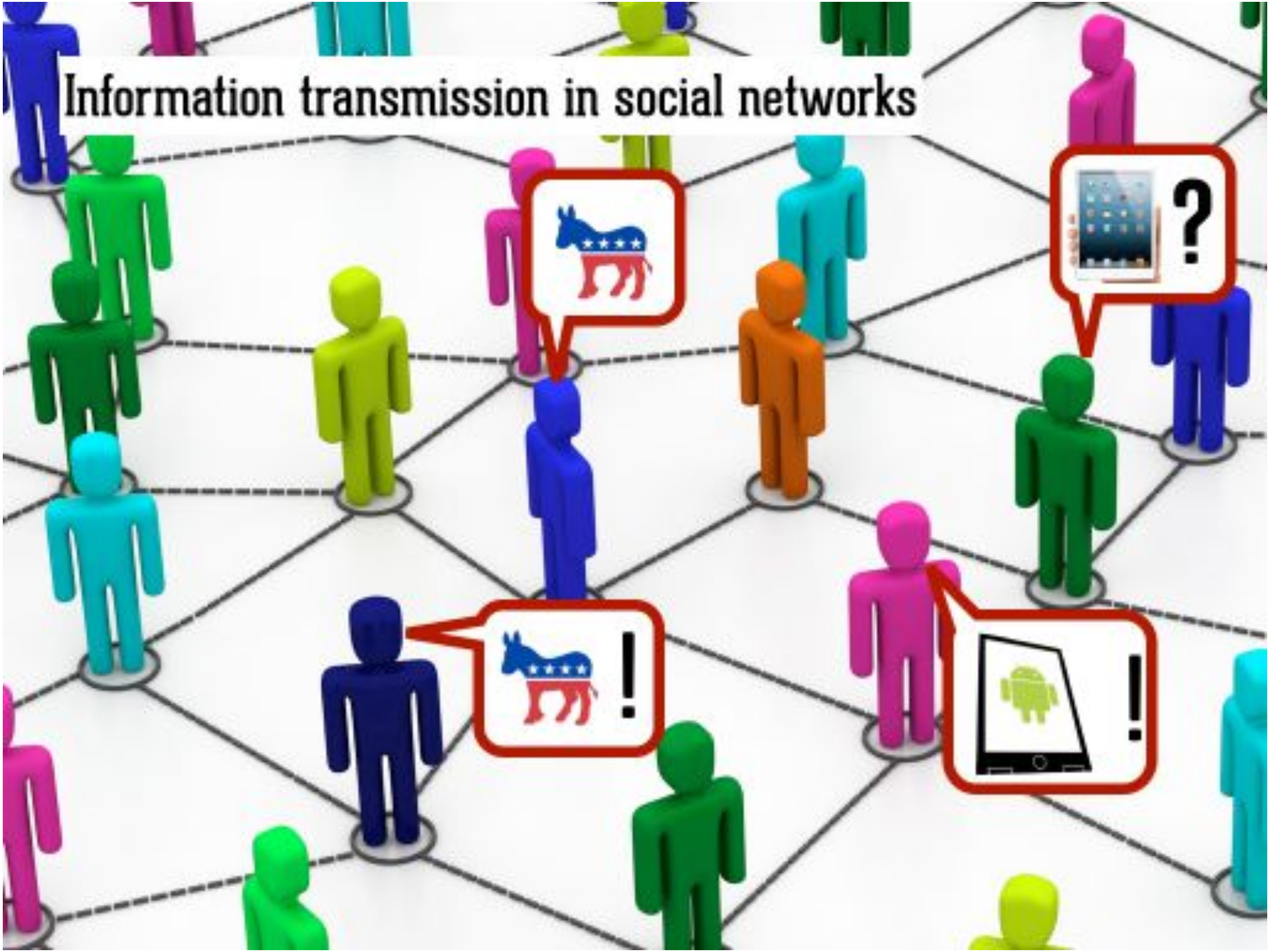

José Luis Iribarren
IIC


Rubén Lara
Telefónica


James Fowler
UCSD

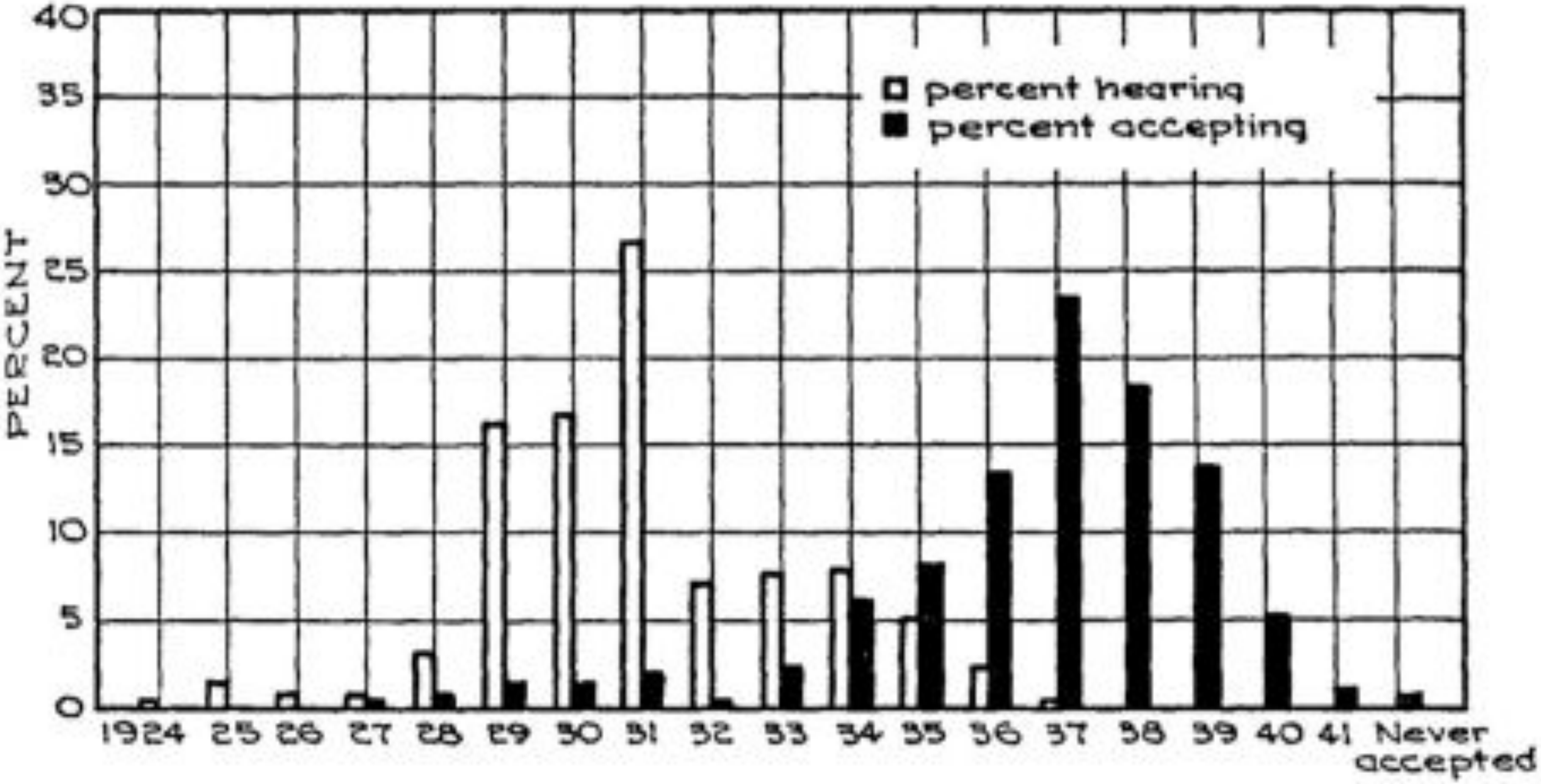

Nicholas Christakis
Yale

@estebanmoro


Information transmission in social networks
?
!
!

# Information transmission on social networks

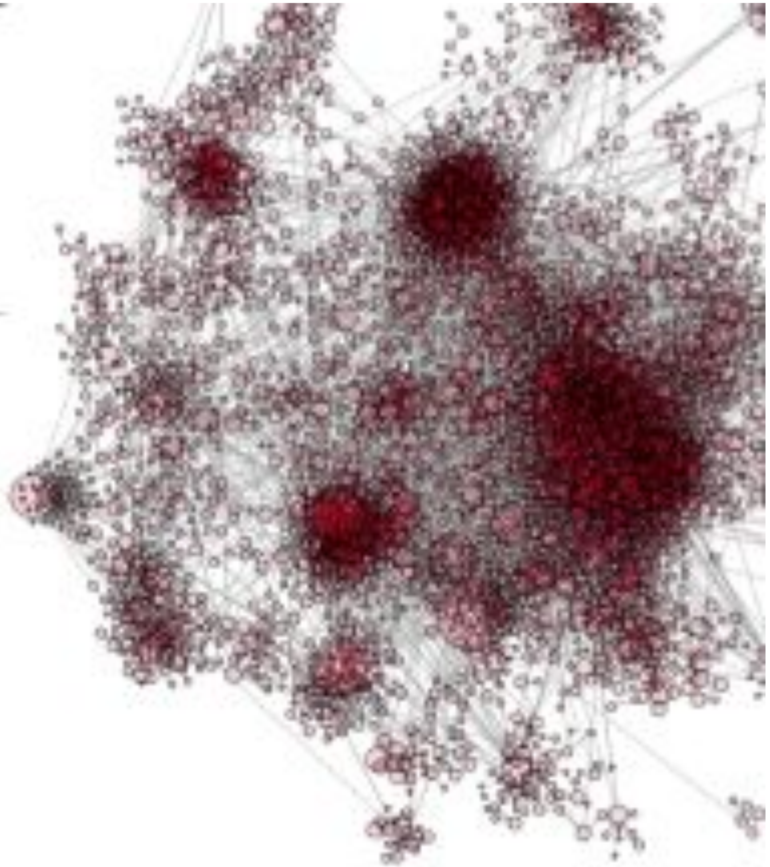
- It plays an essential role in numerous areas
    - Spread of innovations
    - Service/Product adoption
    - Opinion formation
    - Computer viruses
    - Social influence
    - Viral marketing, word of mouth
    - Time-critical emergencies
    - Political communication
    - Social mobilization
    - …




*The Diffusion of Hybrid Seed Corn in Two Iowa Communities '43*

# Relevant questions in spreading

- **Reach**
  - How many people are infected from a initial spreader?
- **Time**
  - How long does it take to infect them?
  - Early detection of an outbreak, possible?
- **Optimization**
  - How do we choose a given a number *N* of initial spreaders, so that reach is maximize in a given time? What is the optimal *N* for a given cost?
  - How do we choose a given number of immune people so that reach of the disease is minimized? (resiliance of networks)
  - How do we choose sensors to detect propagation?



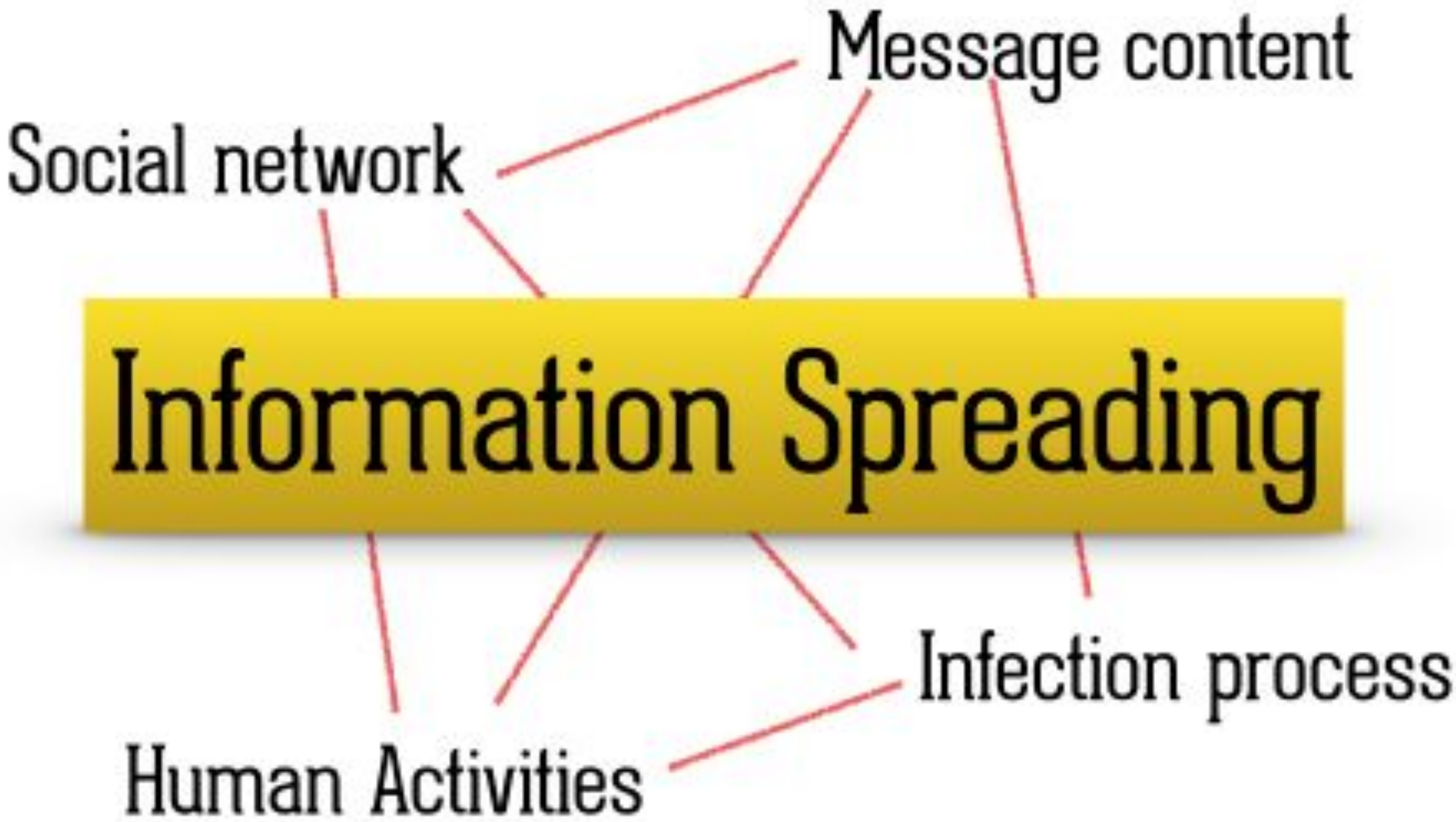
M. García-Herranz, E. Moro, M. Cebrián, J. Fowler & N. Christakis 2012

# What are the ingredients of information spreading

# What are the ingredients of information spreading?

- Social network
- Human activities
- Message content
- Infection process

What are the ingredients of information spreading?



# How do we communicate? Does it impact spreading?

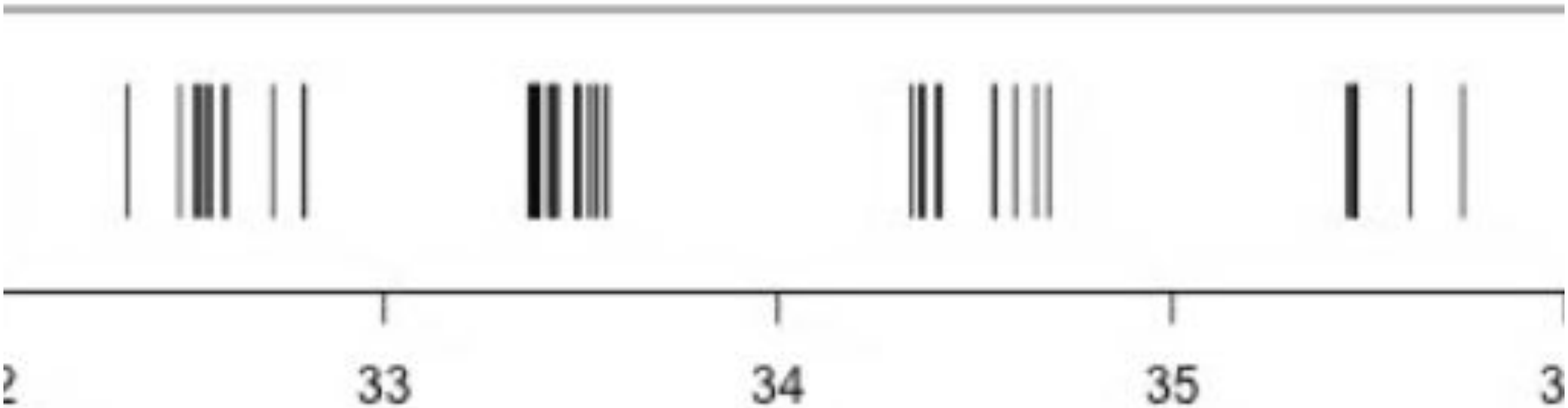
1

iic instituto de ingeniería del conocimiento

Universidad Carlos III de Madrid

@estebanmoro

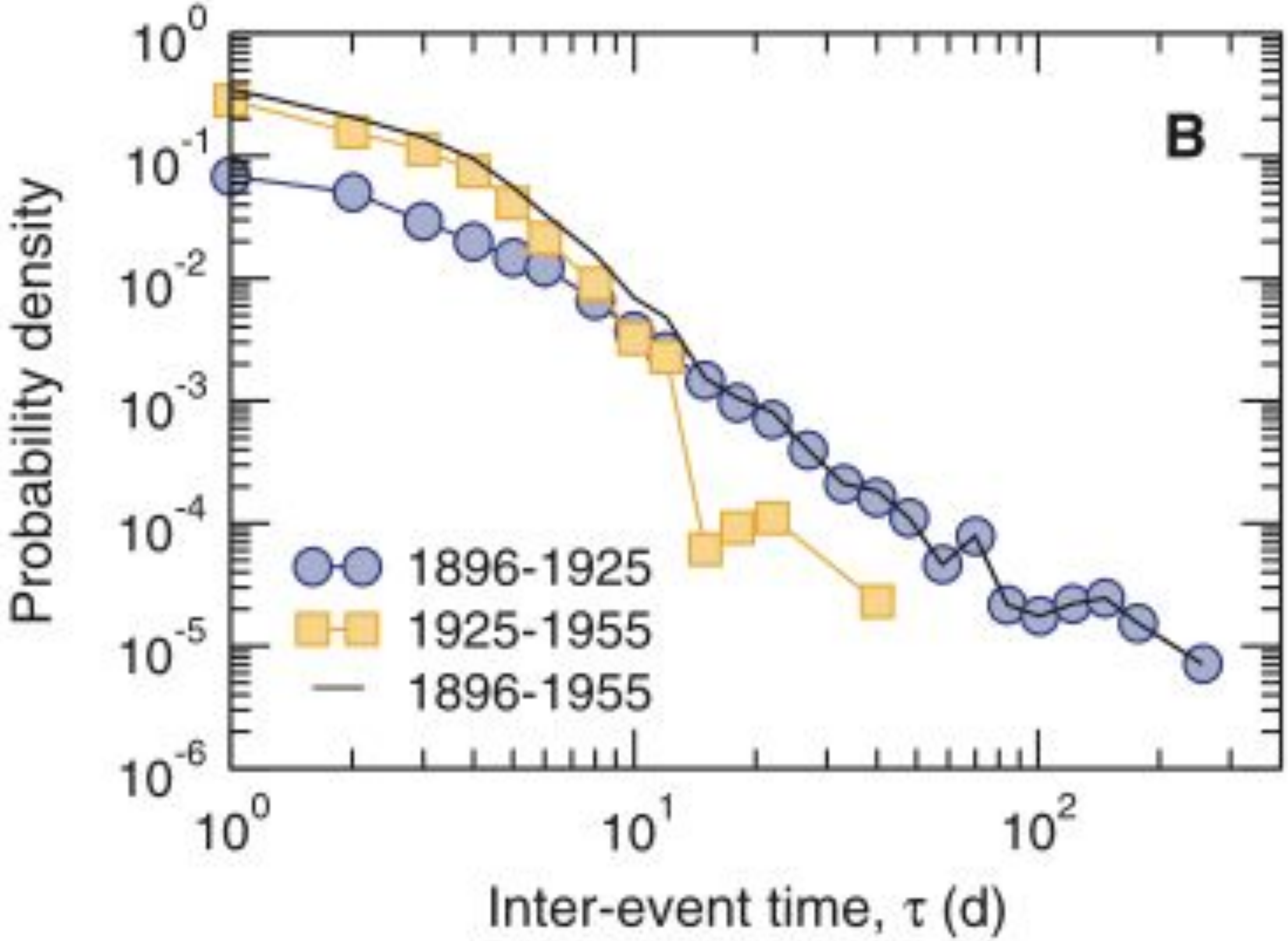
# Humans interact in bursts



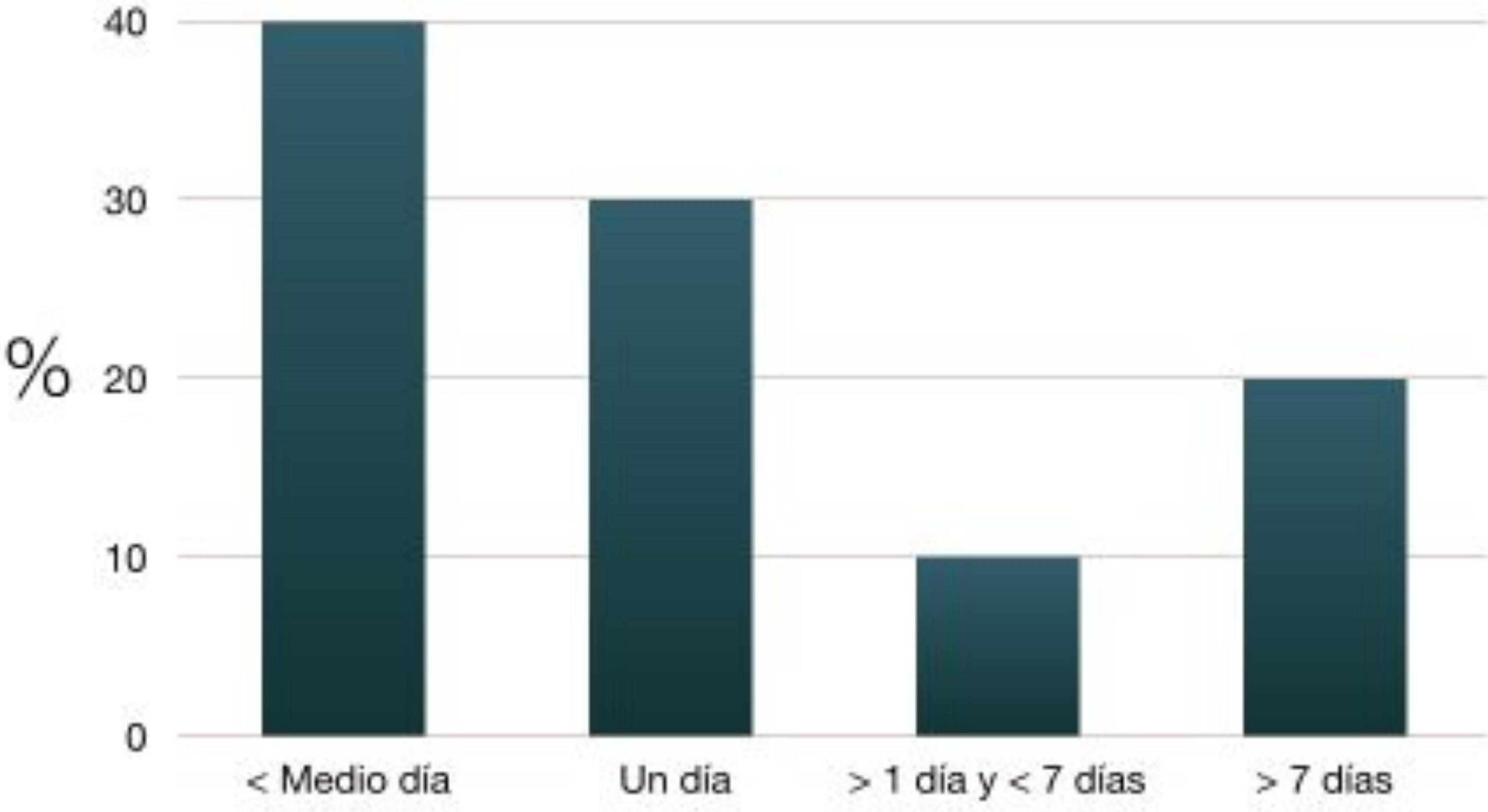

UC3M & Telefónica
Miritello, Moro, Lara 2011/12/13

@estebanmoro

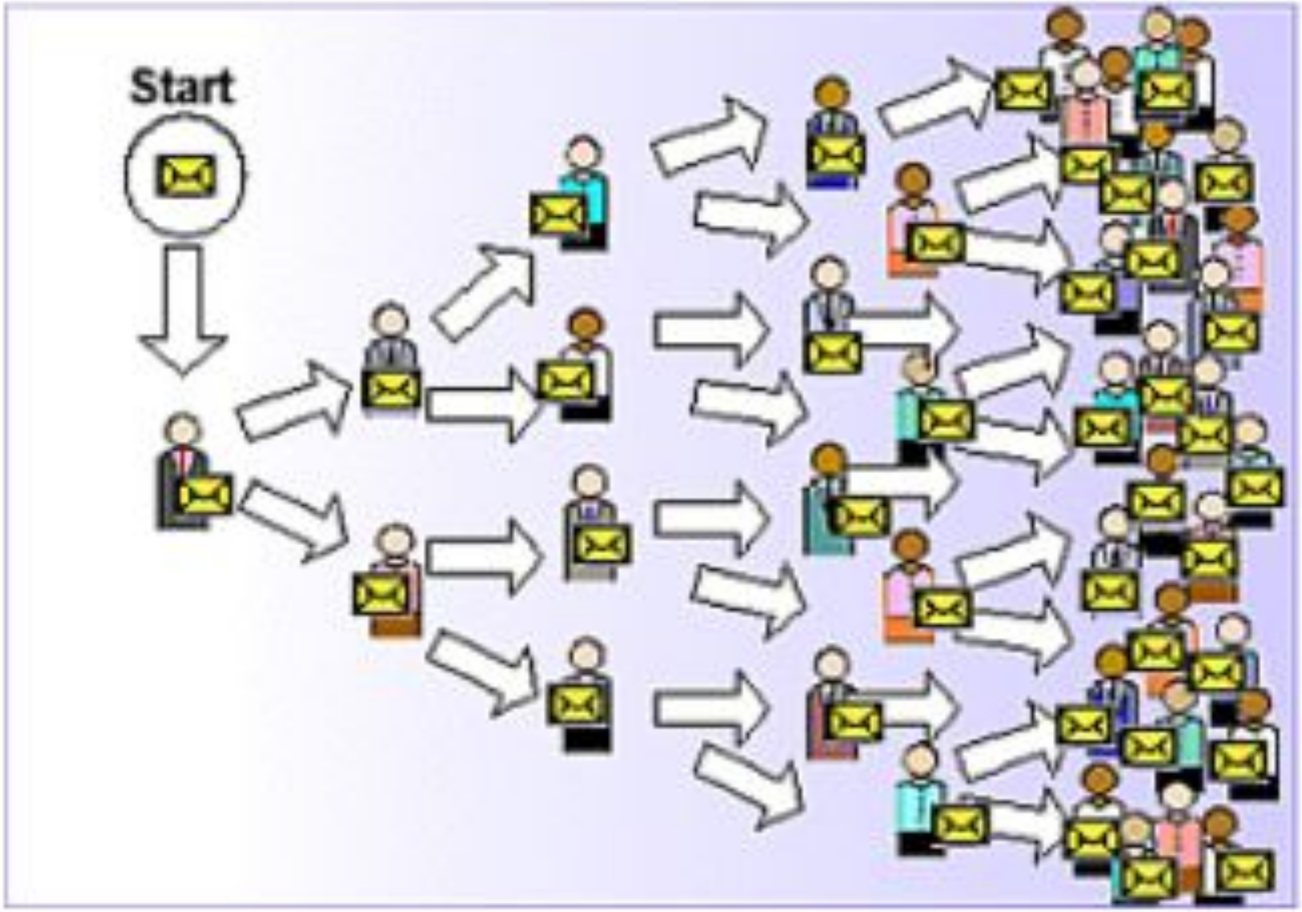
# There is no just a thing as “typical response time”



# There is no just a thing as “typical response time”

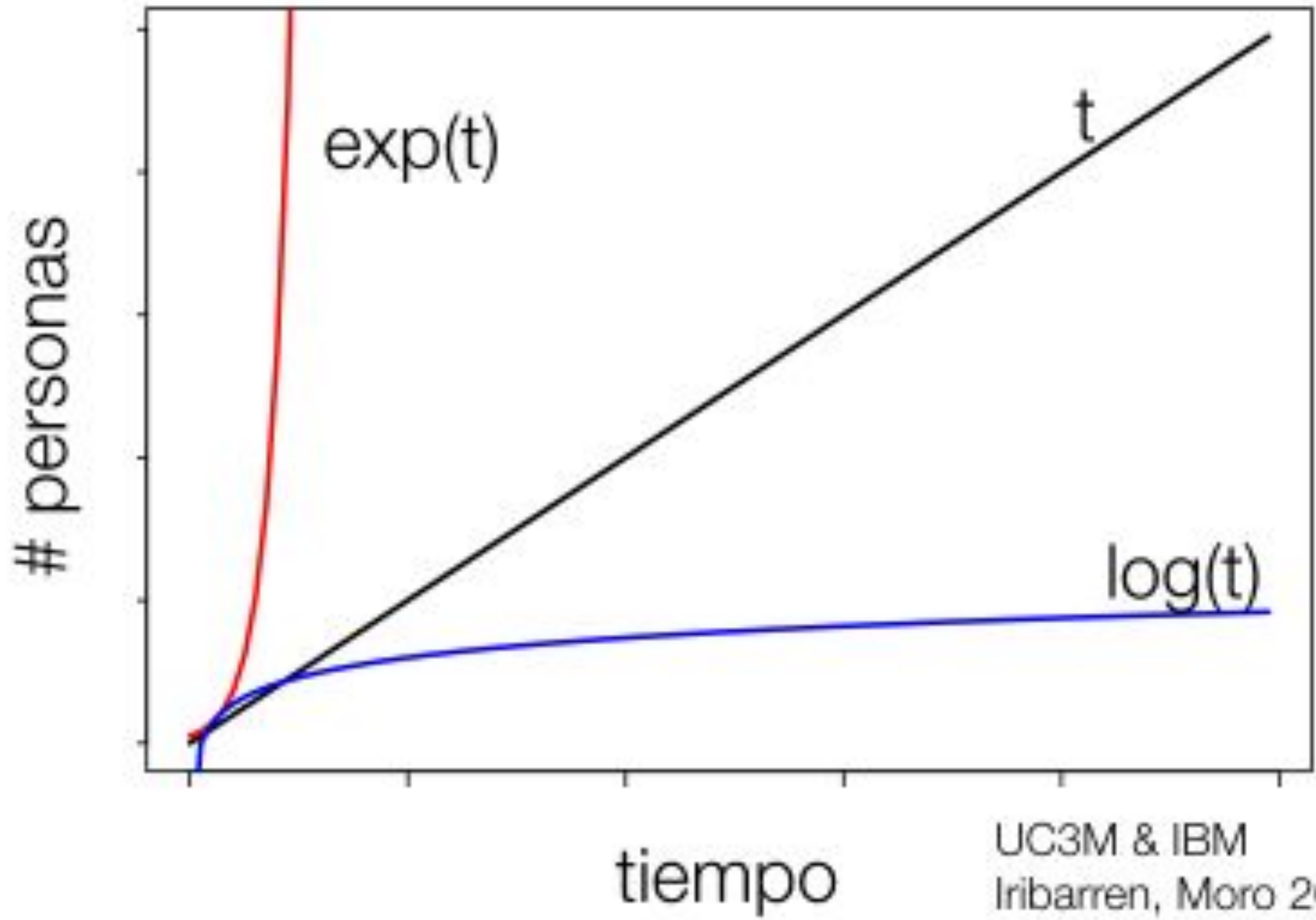


# Thus, what is the "speed of sound" in social networks?




UC3M & IBM
Iribarren, Moro 2009, 2011

# The logarithmic clock of information spreading



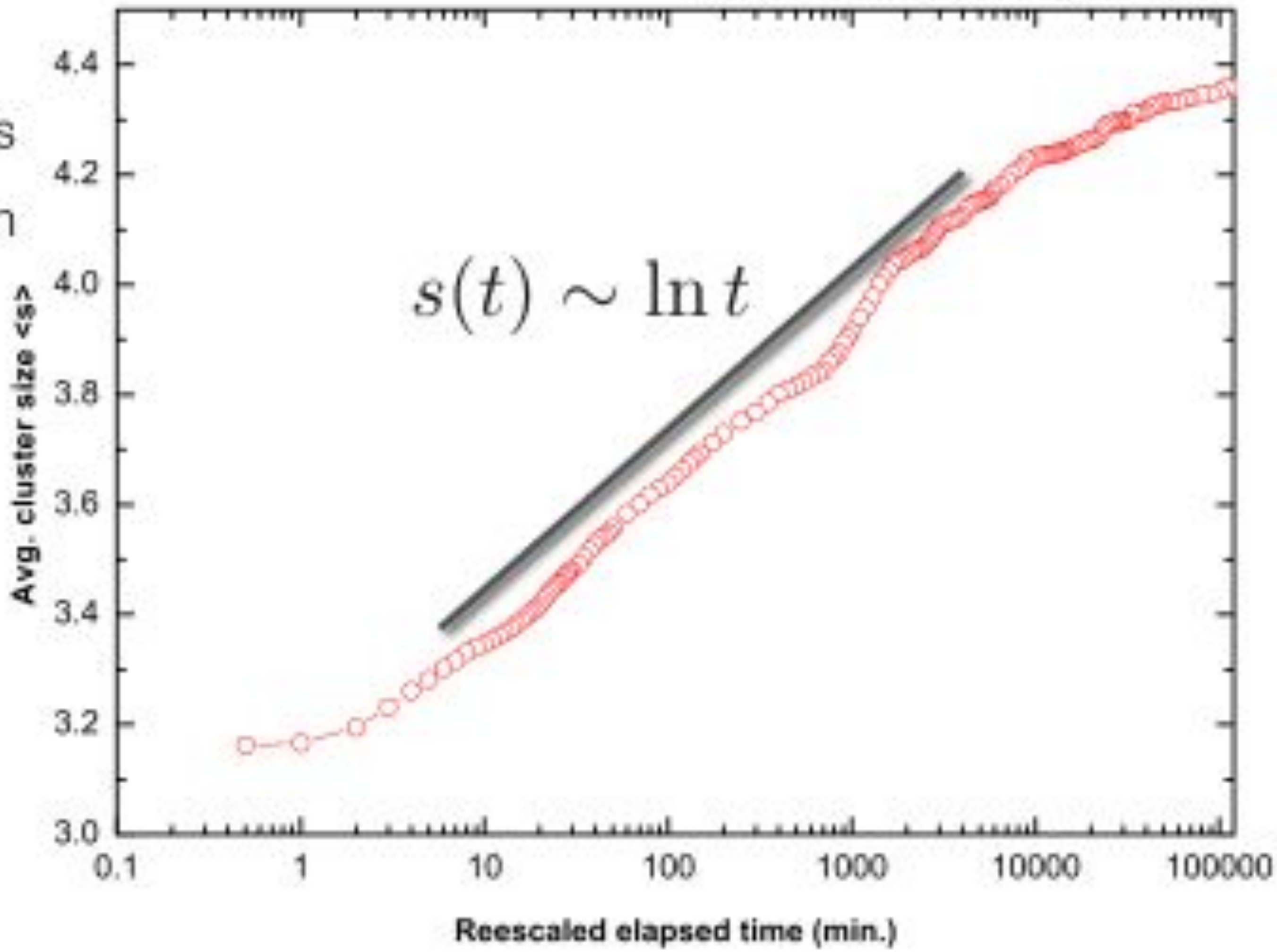

UC3M & IBM
Iribarren, Moro 2009, 2011

@estebanmoro

# Viral marketing campaigns

- 2003-2005 IBM.COM
- 30000 B2B clients
- 11 european countries
- 2 months of campaign

UC3M & IBM
Iribarren, Moro 2009, 2011

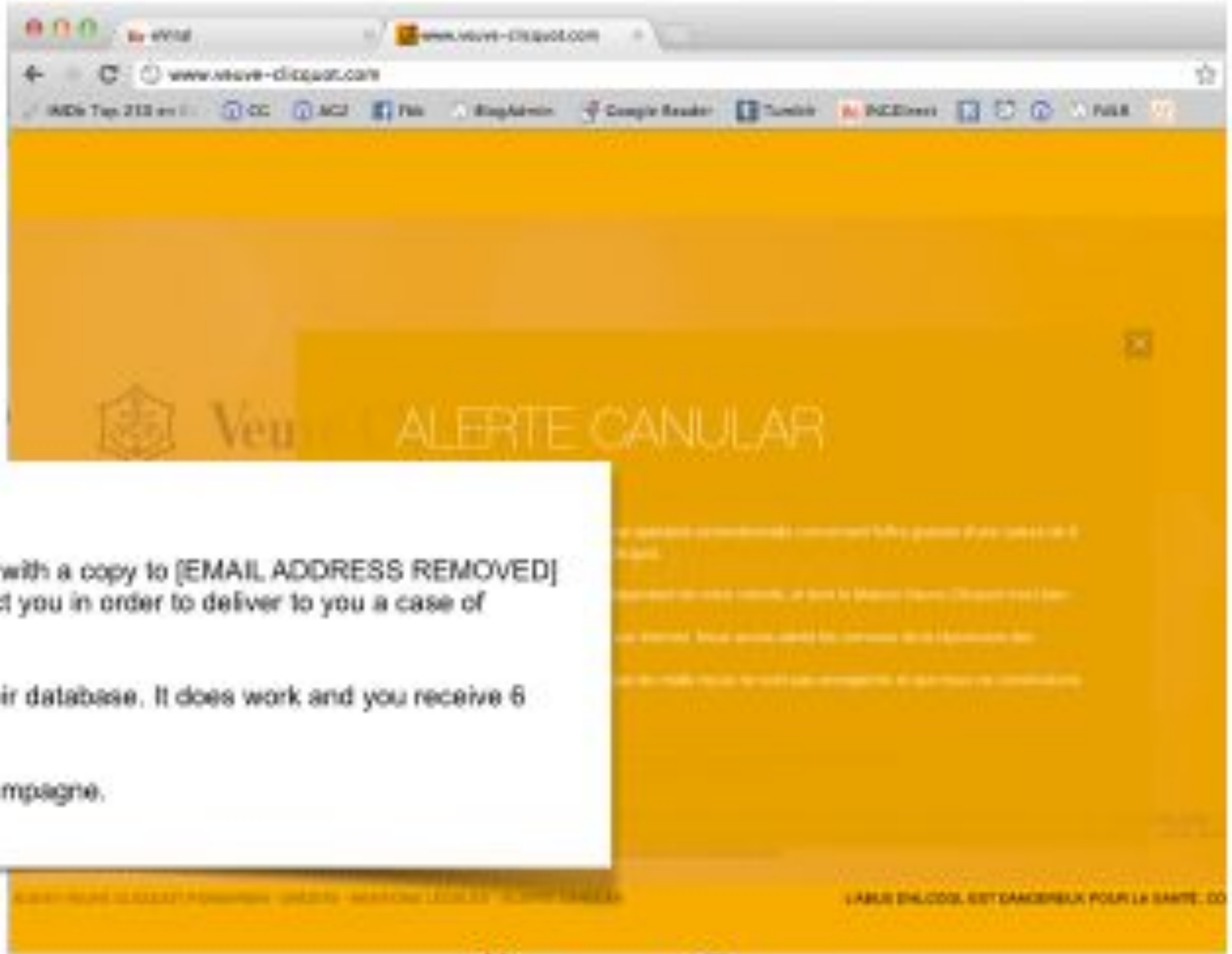



@estebanmoro

# Information prevalence

Average response time = 1 day
Prevalence = 7 days

Average response time = ?
Prevalence = 1 year




Hello all Champagne lovers.

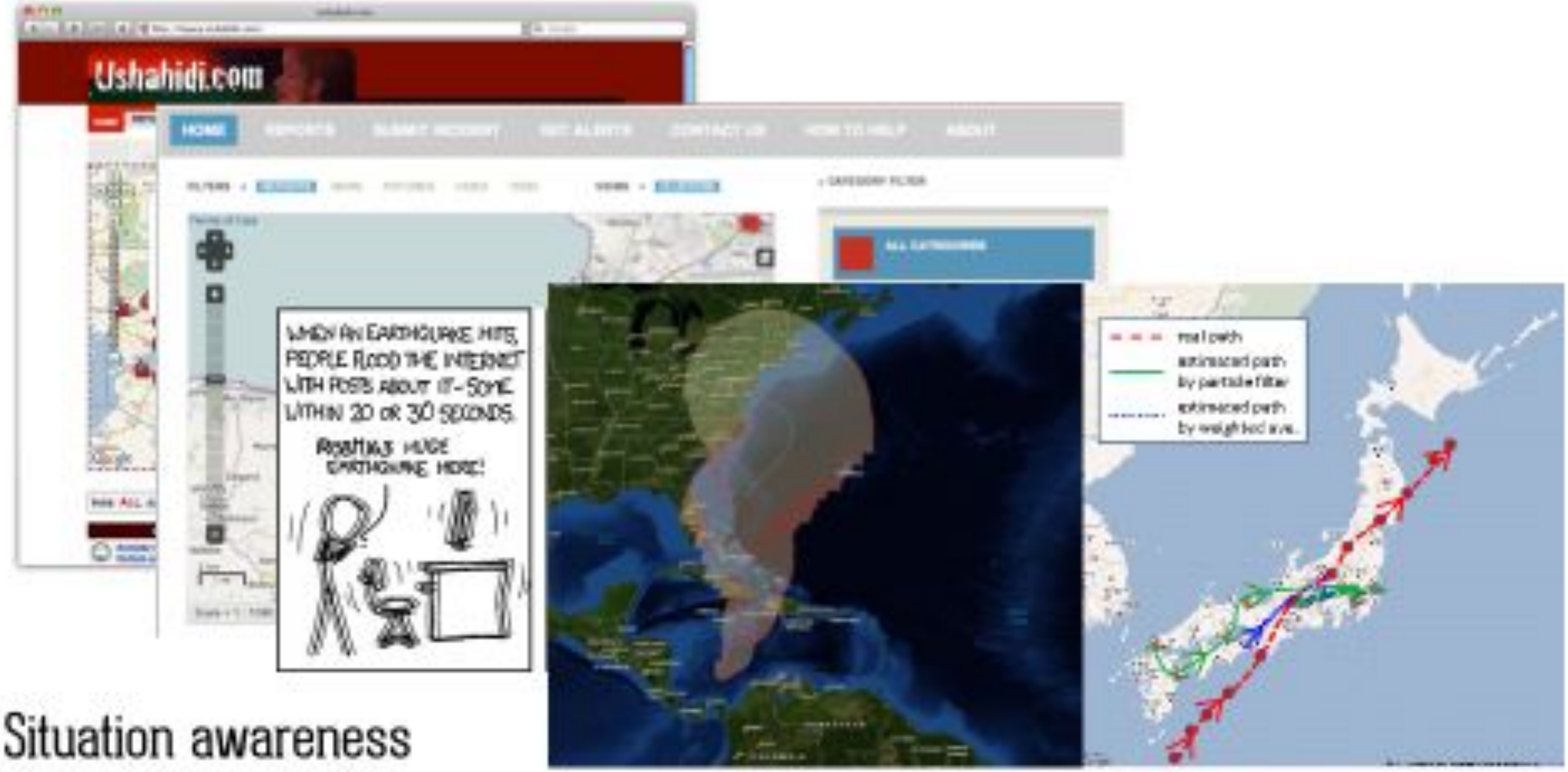
Send this message to 10 people, with a copy to [EMAIL ADDRESS REMOVED] Veuve Clicquot France will contact you in order to deliver to you a case of champagne in three weeks.

They are doing this to enlarge their database. It does work and you receive 6 bottles in 15 days.

Salut a tous les amoureux du champagne.

# 2 How fast and reliable is social mobilization?

# Use social networks to coordinate the mobilization of large masses of people in emergency situations



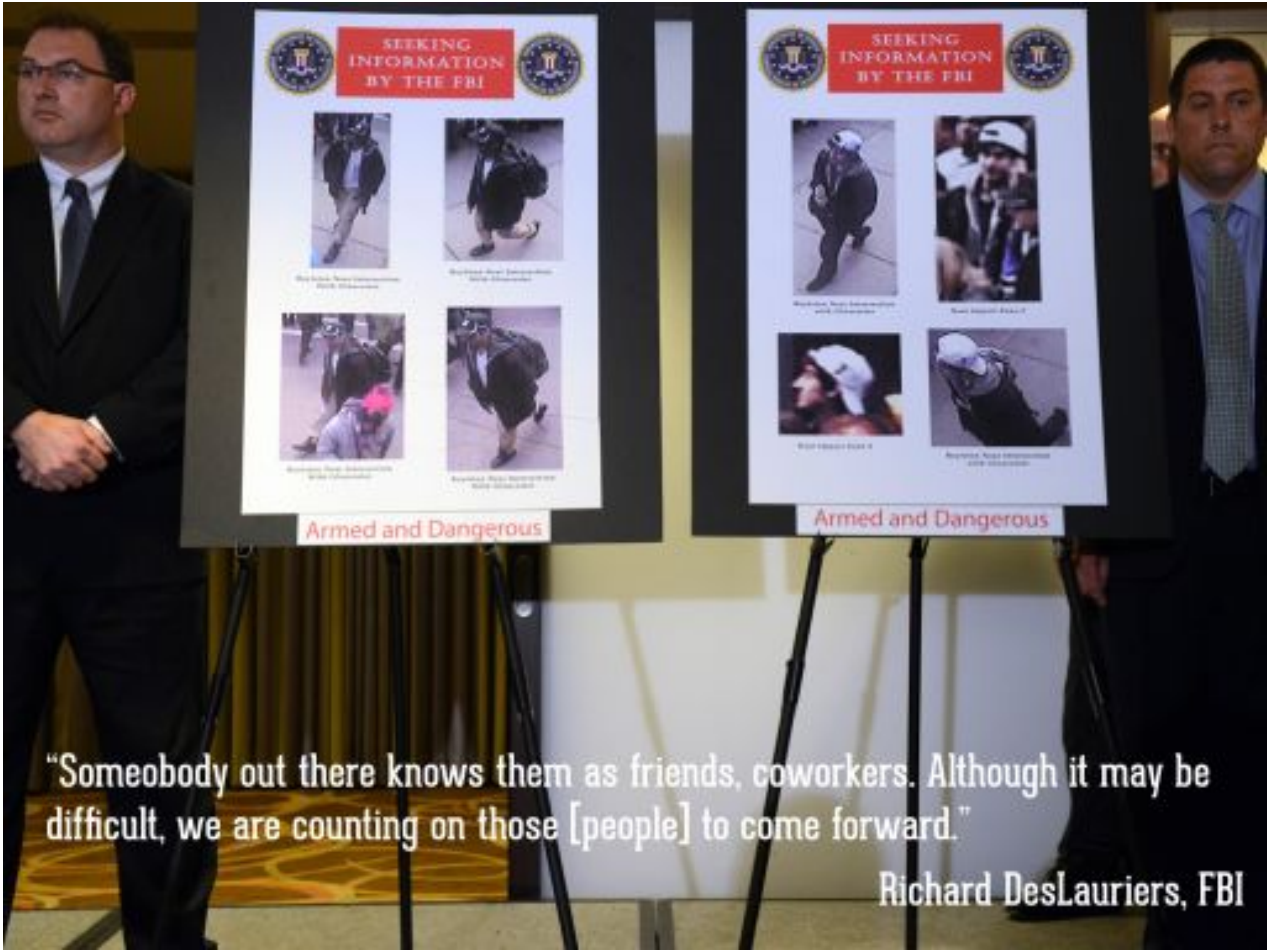

Situation awareness
Crowdsourcing
Event detection
Search/Locating survivors

# Twitter Alerts: Critical information when you need it most

Wednesday, September 25, 2013 | By Gaby Peña (@gpena), Product Manager (16:58 UTC)

Tweet

Today, we're launching Twitter Alerts, a new feature that brings us one step closer to helping users get important and accurate information from credible organizations during emergencies, natural disasters or moments when other communications services aren't accessible.


SEEKING INFORMATION BY THE FBI
SEEKING INFORMATION BY THE FBI
Armed and Dangerous
Armed and Dangerous
“Someobody out there knows them as friends, coworkers. Although it may be difficult, we are counting on those [people] to come forward.”
Richard DesLauriers, FBI



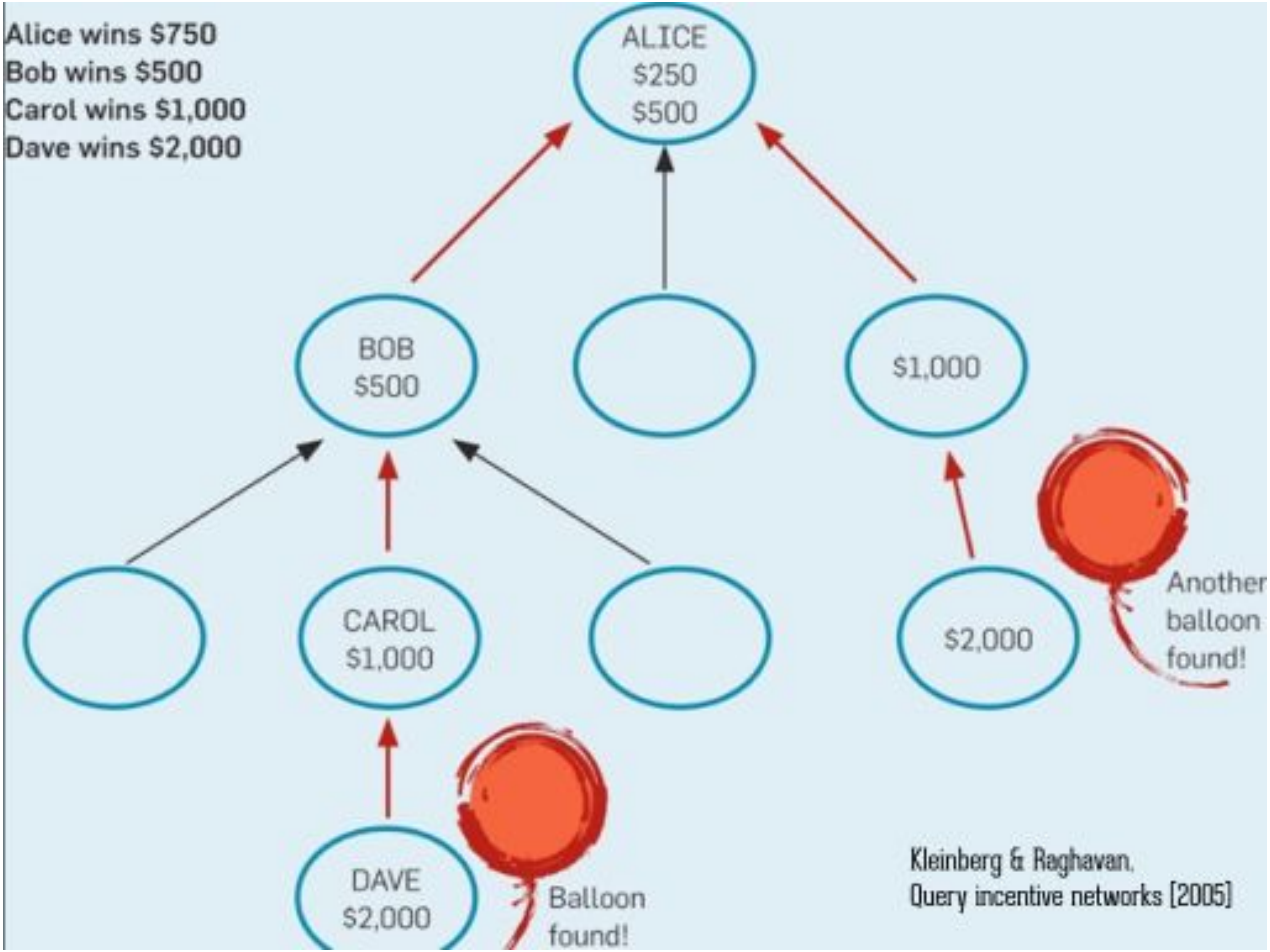

"Impossible by
conventional intelligence"

@estebanmoro

# Time-Critical Social Mobilization

Galen Pickard,[1,2*] Wei Pan,[1*] Iyad Rahwan,[1,3*] Manuel Cebrian,[1*] Riley Crane,[1]
Anmol Madan,[1] Alex Pentland[1†]

The World Wide Web is commonly seen as a platform that can harness the collective abilities of large numbers of people to accomplish tasks with unprecedented speed, accuracy, and scale. To explore the Web's ability for social mobilization, the Defense Advanced Research Projects Agency (DARPA) held the DARPA Network Challenge, in which competing teams were asked to locate 10 red weather balloons placed at locations around the continental United States. Using a recursive incentive mechanism that both spread information about the task and incentivized individuals to act, our team was able to find all 10 balloons in less than 9 hours, thus winning the Challenge. We analyzed the theoretical and practical properties of this mechanism and compared it with other approaches.


Alice wins $750
Bob wins $500
Carol wins $1,000
Dave wins $2,000
ALICE
$250
$500
BOB
$500
$1,000
CAROL
$1,000
$2,000
Another balloon found!
DAVE
$2,000
Balloon found!
Kleinberg & Raghavan,
Query incentive networks [2005]

# Global reach in **36 hours**



# A model of social geographical mobilization

PNAS

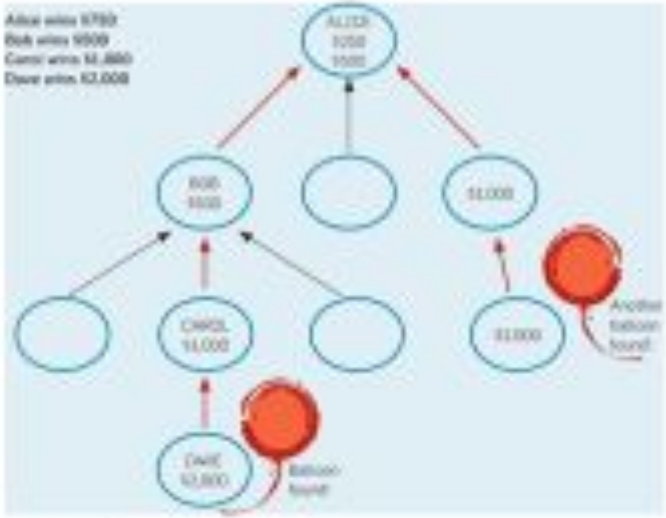
## Limits of social mobilization

Alex Rutherford[a], Manuel Cebrian[b,c], Sohan Dsouza[a], Esteban Moro[d,e], Alex Pentland[f], and Iyad Rahwan[a,g,1]

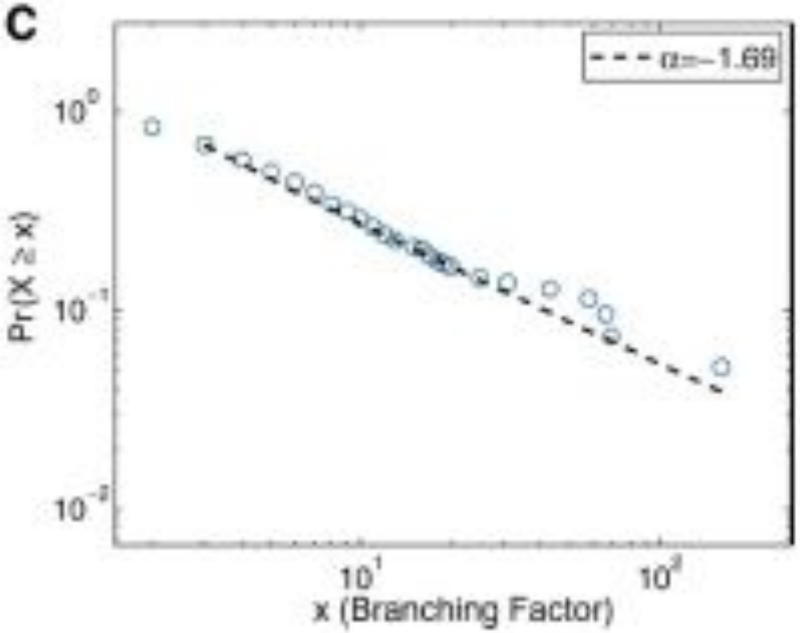
[a]Computing and Information Science, Masdar Institute of Science and Technology, Abu Dhabi 54224, United Arab Emirates; [b]Department of Computer Science and Engineering, University of California at San Diego, La Jolla, CA 92093; [c]National Information and Communications Technology Australia, Melbourne, VIC 3010, Australia; [d]Departamento de Matemáticas and Grupo Interdisciplinar de Sistemas Complejos, Universidad Carlos III de Madrid, 28911 Madrid, Spain; [e]Instituto de Ingeniería del Conocimiento, Universidad Autónoma de Madrid, 28049 Madrid, Spain; [f]Media Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; and [g]School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom


:tebanmoro

# Branching dynamics of recruitment



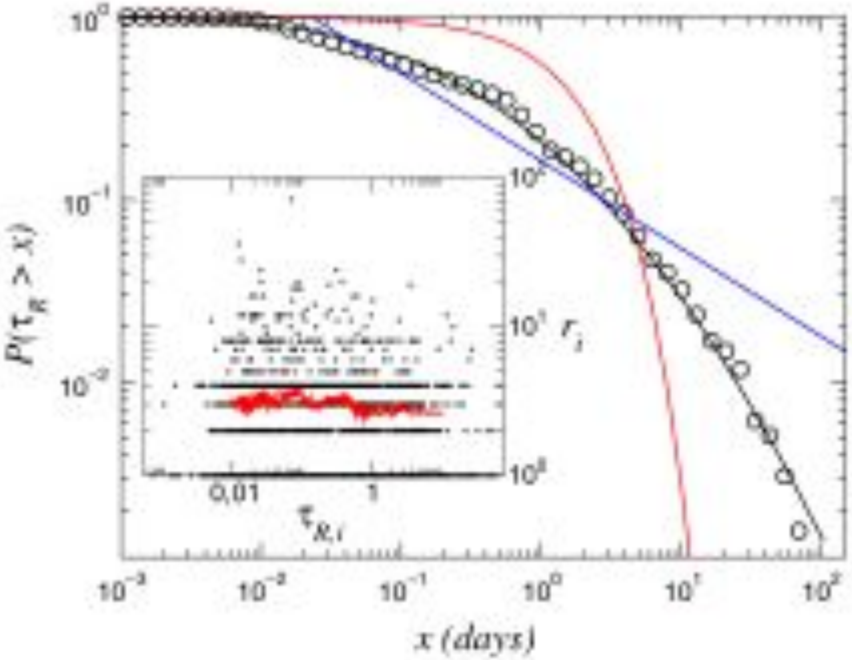
Number of recruitments is power law distributed
Below tipping point $\langle R_0 \rangle = 0.89$
Only 4400 active recruitments
We consider also passive recruitments [100s per person]




# Temporal dynamics of message propagation

Iribarren JL and Moro E (2009) PRL
Impact of human activity patterns on Dynamics of Information Diffusion

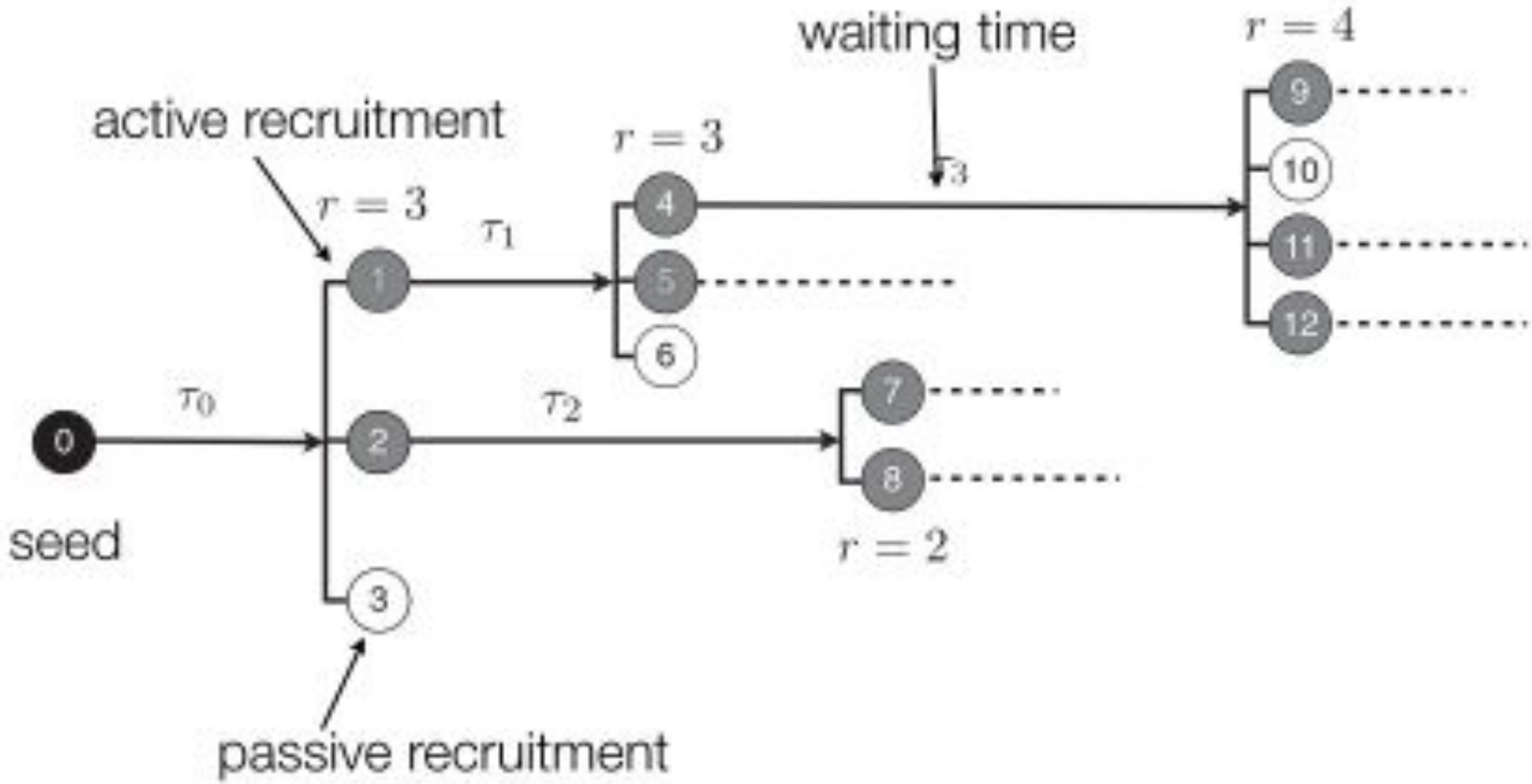
Waiting time between receiving invitation and sending it is log-normal distributed $[\mu = 1.5, \sigma = 5.5]$
Above the tipping point diffusion is dominated by short waiting times



# Viral dynamics of recruitment $\approx$ viral marketing

Branching dynamics of recruitment

Temporal dynamics of message propagation



# Geographical spreading of social networks

## Geographical scaling laws of friendship

**Gravity Law**

Krings et al JSTAT 2009 / Ratti et al PLoS ONE 2010

$$P_{ij} \propto \frac{1}{(d_{ij})^{\alpha}}$$

**Rank-based friendship**

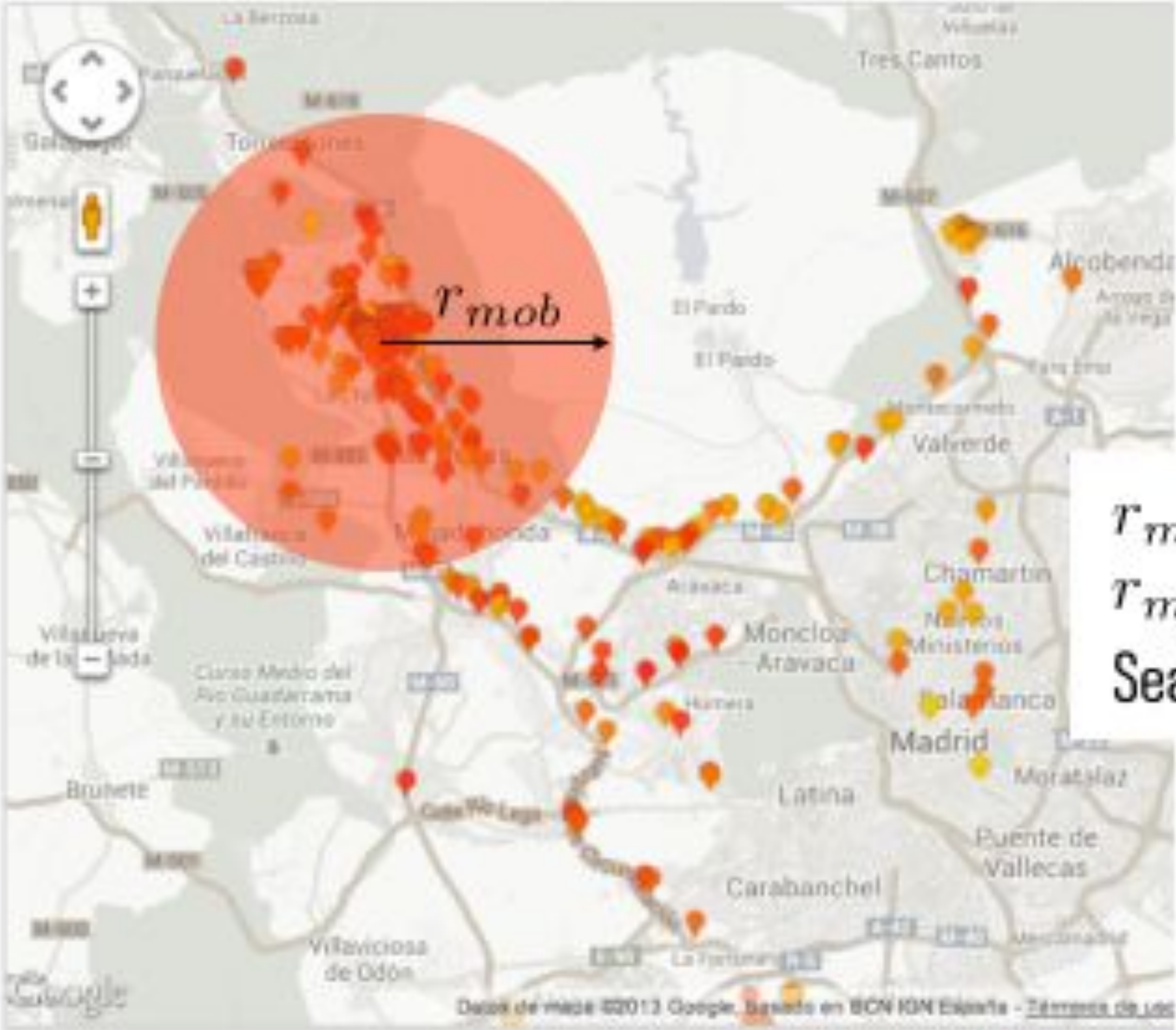
Liben-Nowell D et al PNAS 2005

$$P_{ij} \propto \begin{cases} \frac{1}{\sum_{k:r_{ik}<r_{ij}} p_k} & \text{with probability } 2/3 \\ 1 & \text{with probability } 2/3 \end{cases}$$

Two people separated by a given large distance are more likely to be friends in a rural region than in a dense, urban environment

# Human mobility

Due to mobility, searching is not limited to people's home

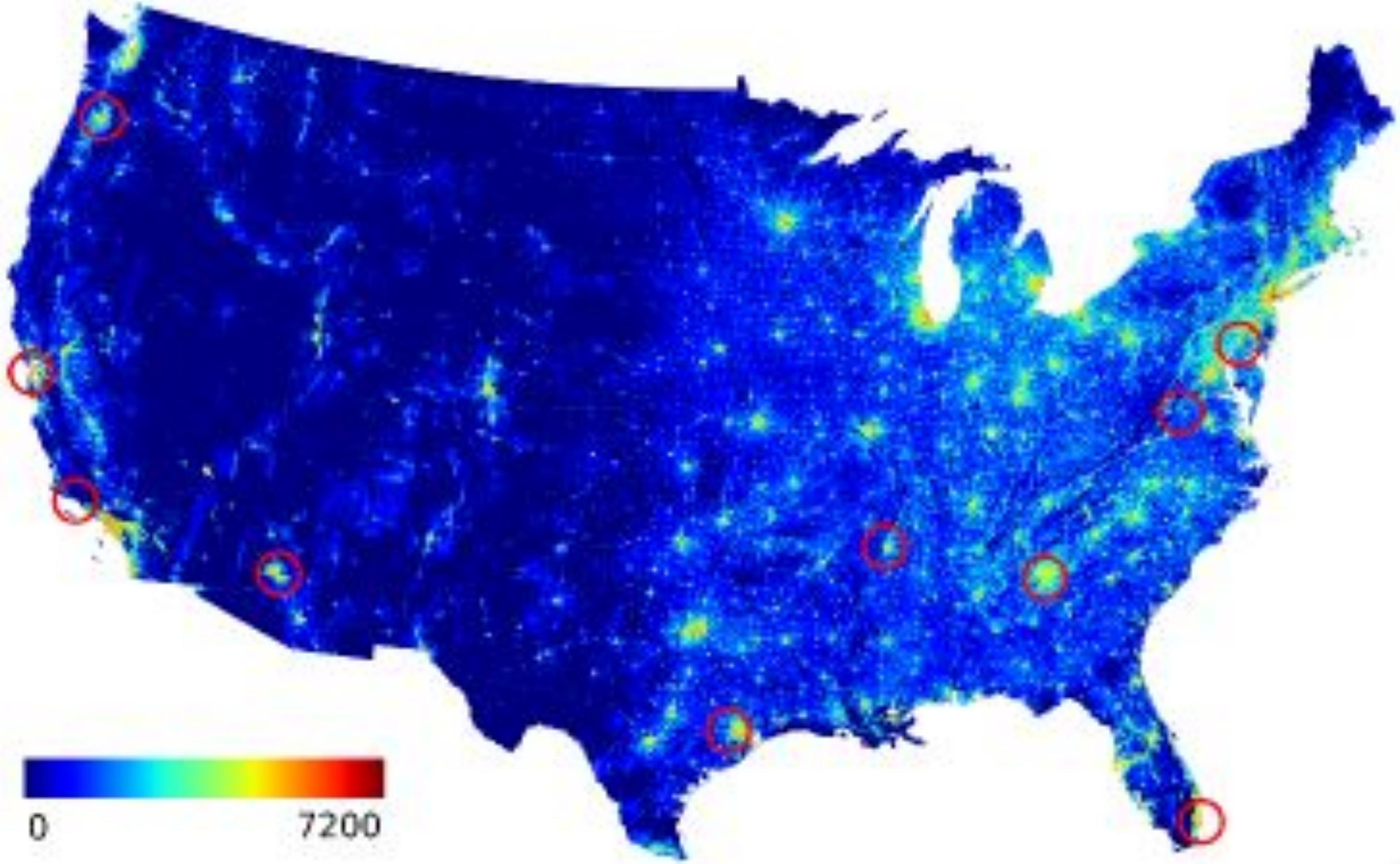



$r_{mob} \simeq 1 - 2km\ (12h)$
$r_{mob} \sim 1/p_k$
Search area $\simeq (r_{mob})^2$

@estebanmoro

# Census data

7,820,528 1km² cells
2,760,240 unoccupied
Heterogeneity and clustering



# Model

1. Select a seed [@MIT]
2. Wait for a response time
3. Recruit a number of active/passive new members
4. Choose them on short/large distances
5. If ballon is in the search area of the recruit => found
6. Proceed to 2 with the active recruits. If none, stop

## Results

We conducted **500 searches** for the 10 DARPA Network Challenge

**Parameters**:

Radius of gyration [mobility]

Number of passive recruitments [viral recruitment]

We **measured**

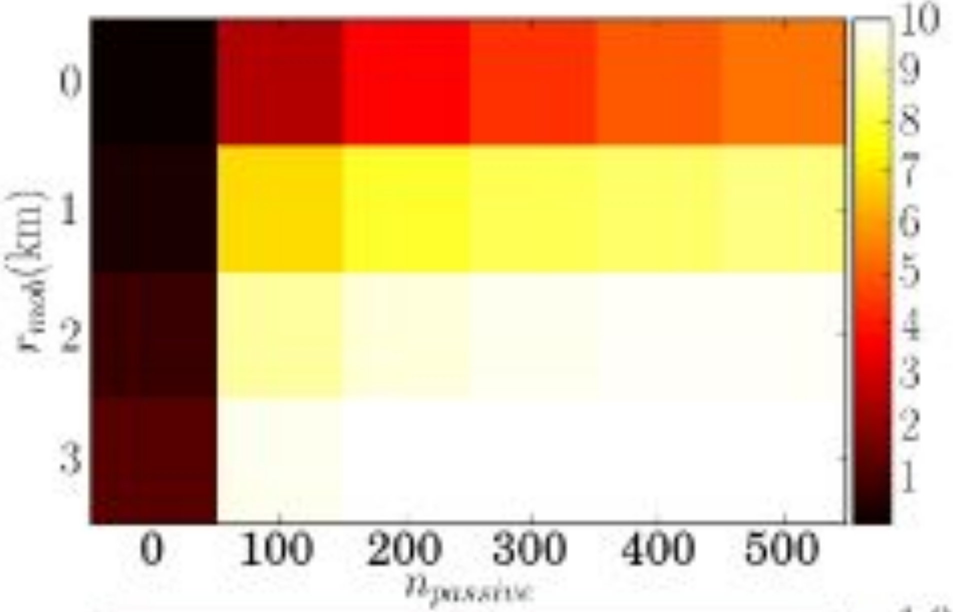
Number of balloons located

Probability of success

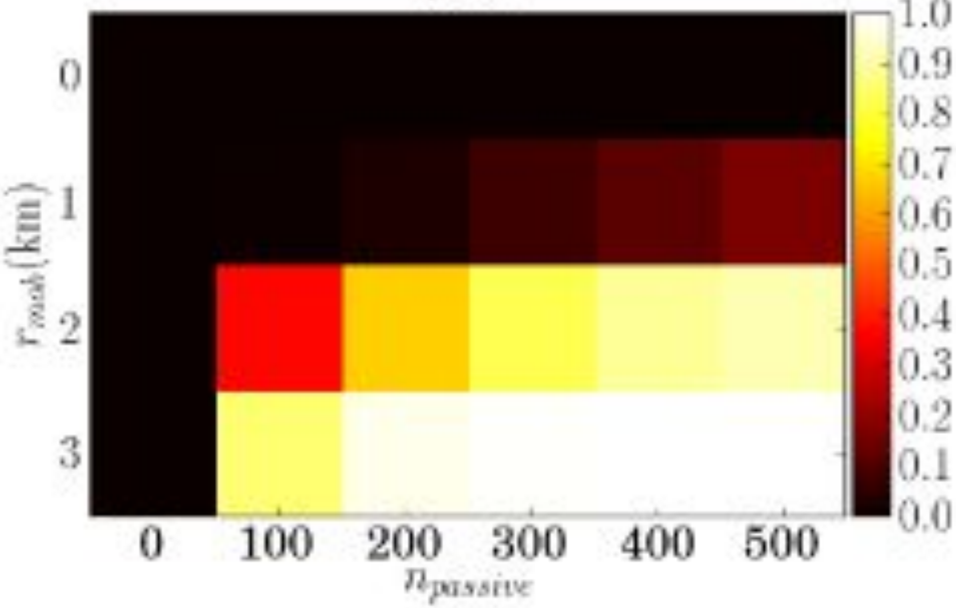
Time to completion

# Results

Average number of balloons found

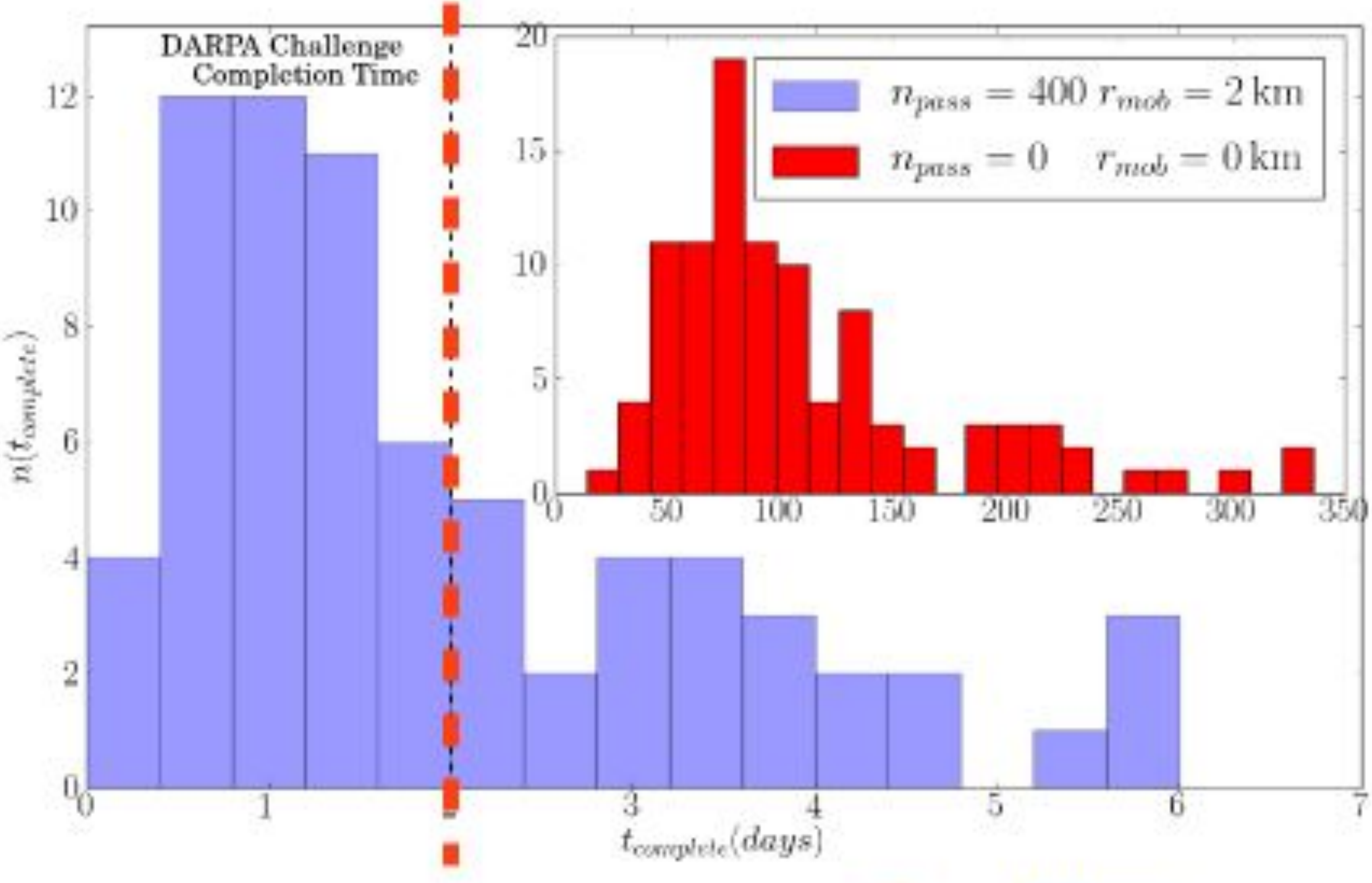



Probability of 100% success



# Results

## Time to completion




@estebanmoro

# Outlook

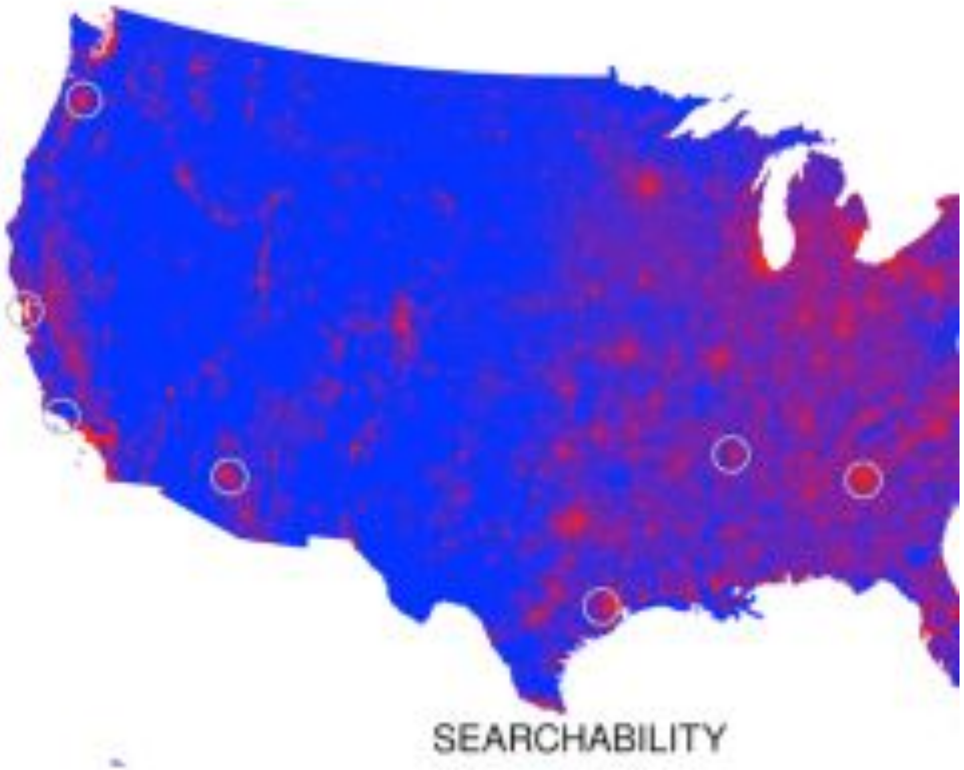
Social mobilization in critical time **can** work

But success relies on:

- Incentives to **participate**/search
- Incentives to **recruit**
- Use of **geographical heuristics**

Challenges

- Urban **complexity**
- **Verification** of information
- **Data-driven** simulations



SEARCHABILITY

@estebanmoro

# 3 Can we detect information outbreaks in social networks?

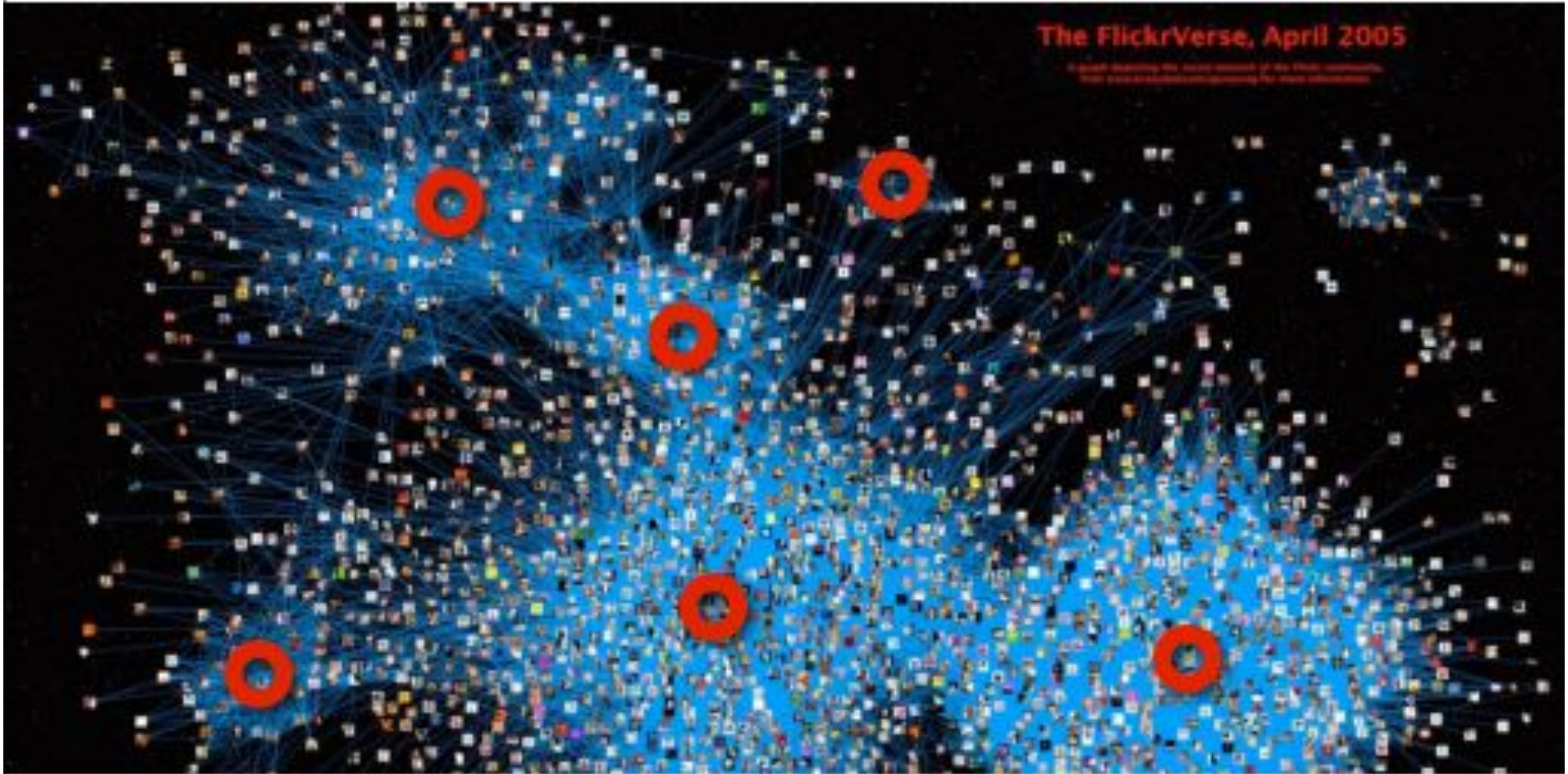
# **Outbreak detection** in networks


C. Nicolaides et al, '12

# **Outbreak detection** in networks


C. Nicolaides et al, '12

# Sensor hypothesis

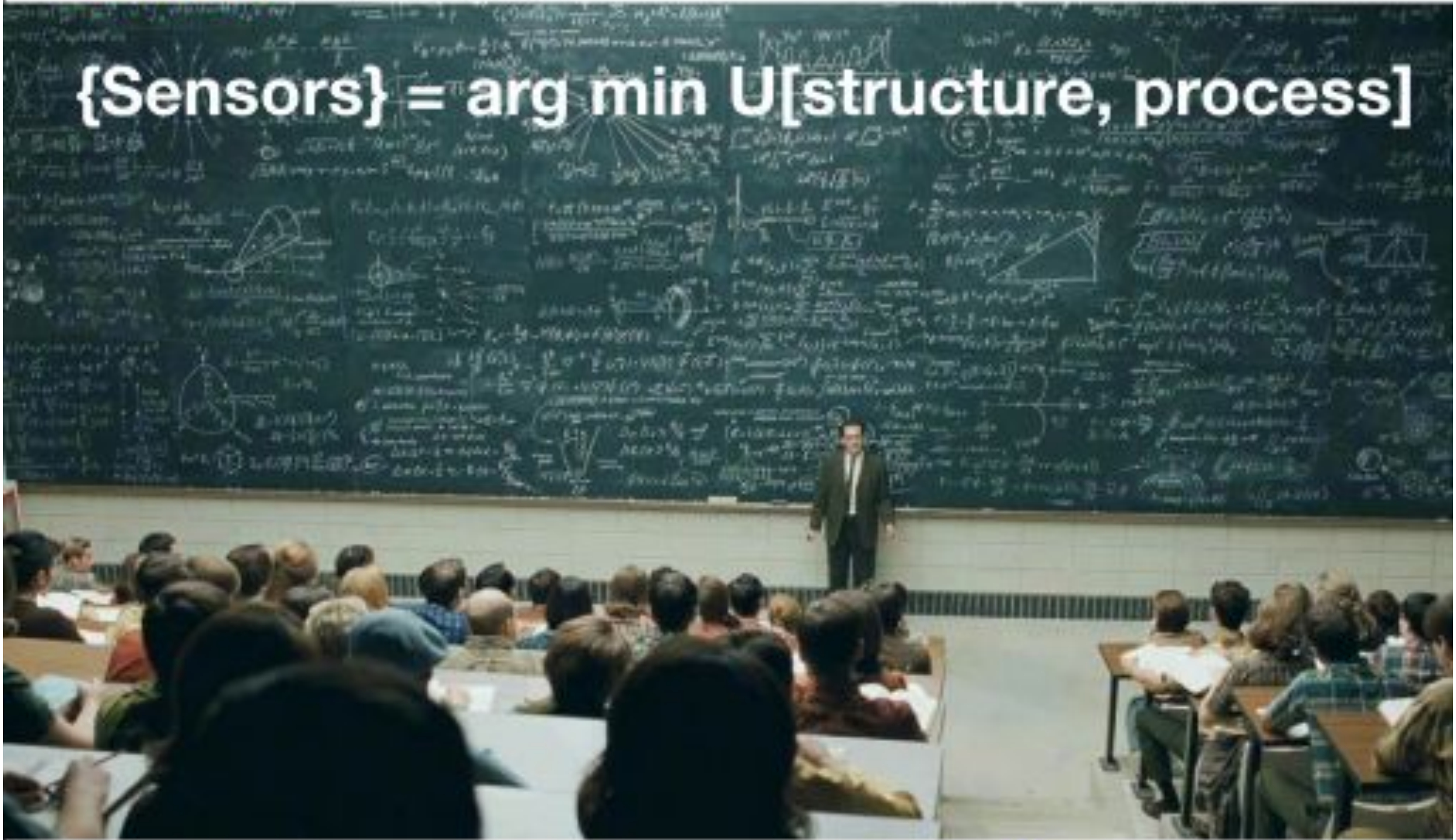



**Define a set of nodes (sensors) to detect outbreaks**

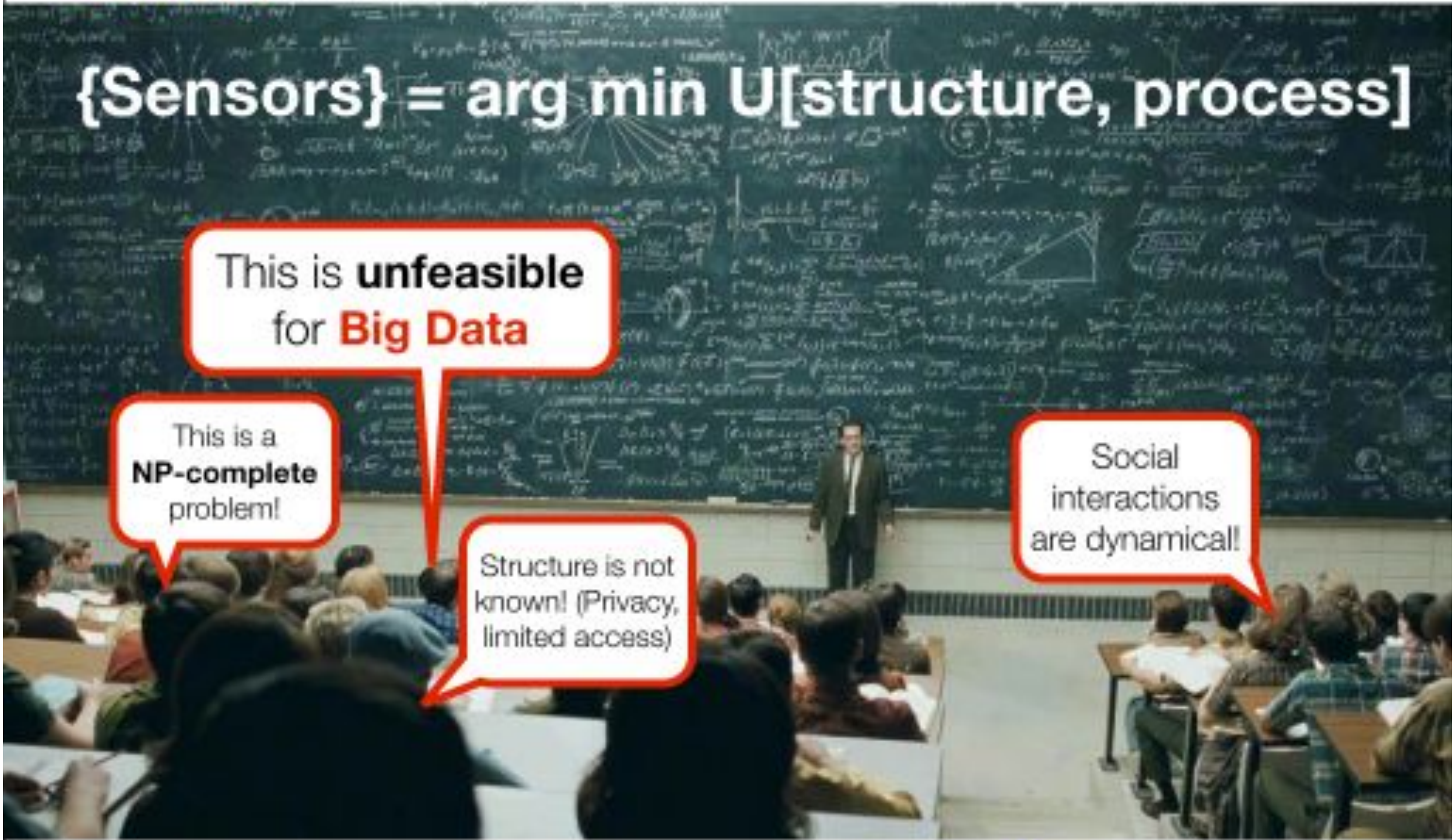
**Can we find those sensors to early detect planetary information outbreaks?**

Universidad Carlos III de Madrid @estebanmoro

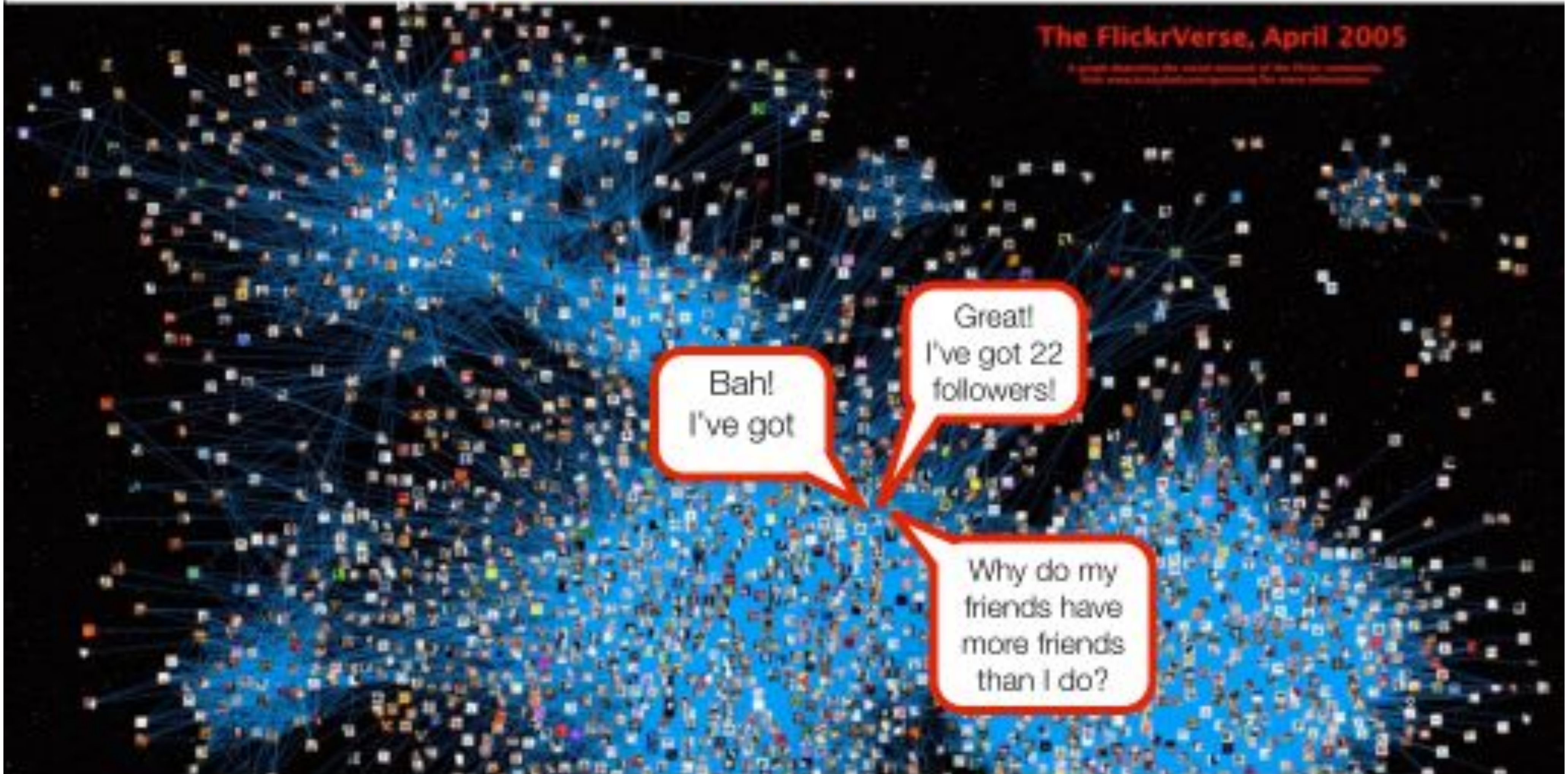
## **Sensor** hypothesis



# Sensor hypothesis



# Our solution: FriendSensors




**Use the friendship paradox (Feld): the friends of a set of users are more central**

# Our solution: **friendsensors**


D. Hansen, B. Shneiderman, M. Smith, Analyzing Social Media Networks with NodeXL: Insights from a Connected World

$$\langle t_{inf,i} \rangle_{i \in \mathrm{Sensors}} \leq \langle t_{inf,i} \rangle_{i \in \mathrm{Control}}$$

# Our solution: FriendSensors

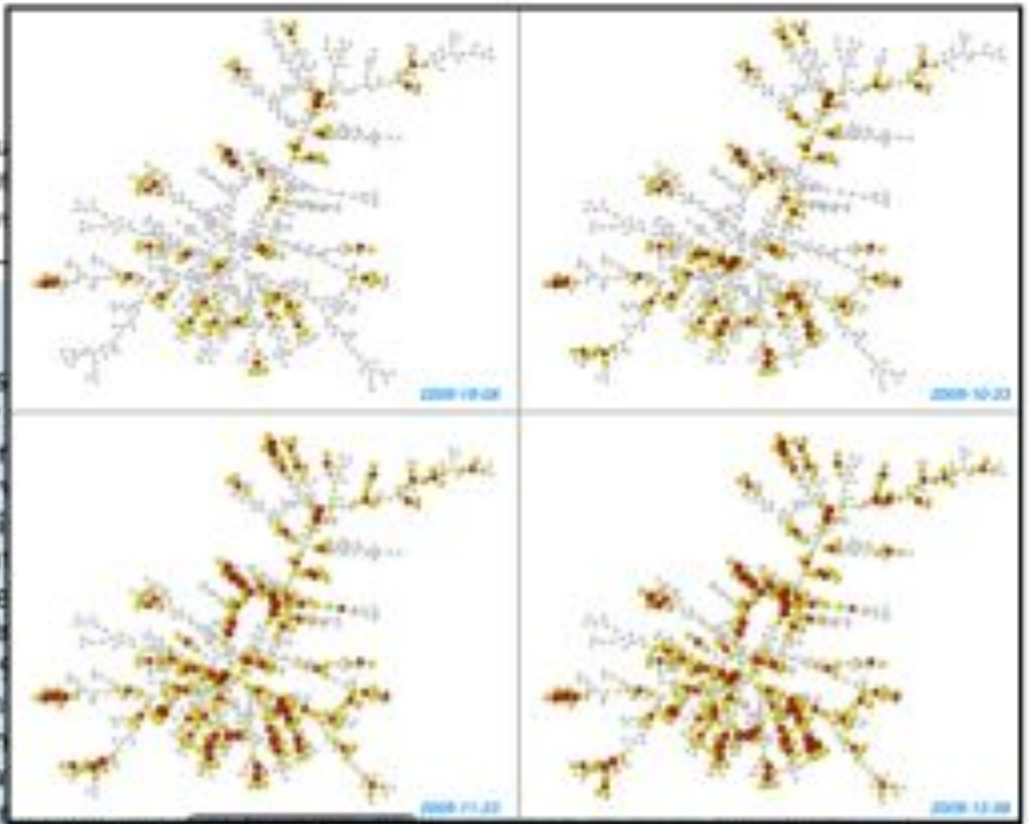
OPEN ACCESS Freely available online

PLoS one

## Social Network Sensors for Early Detection of Contagious Outbreaks

**Nicholas A. Christakis[1,2*], James H. Fowler[3,4]**

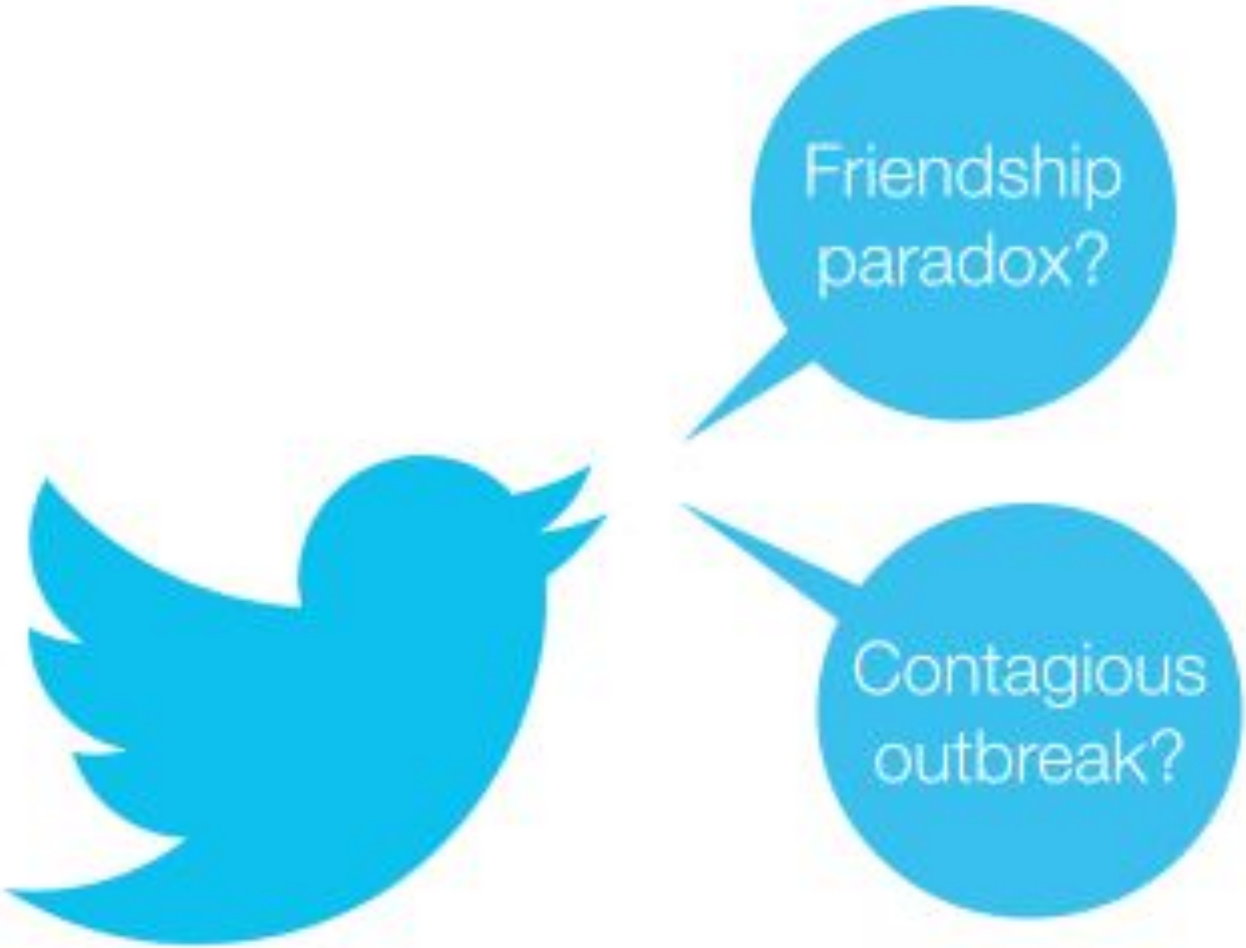
1 Faculty of Arts & Sciences, Harvard University, Boston, Massachusetts, United St
Massachusetts, United States of America, 3 School of Medicine, University of Cali
Sciences, University of California San Diego, La Jolla, California, United States of A



**Abstract**

Current methods for the detection of contagious outbreaks
epidemic at best. It is known that individuals near the center
course of an outbreak, on average, than those at the periphe
individuals who might be monitored for infection is typically
require ascertainment of global network structure, namely, s
Such individuals are known to be more central. To evalua
detection, we studied a flu outbreak at Harvard College in late
a group of randomly chosen individuals or a group of thei
epidemic in the friend group occurred 13.9 days (95% C.I.
population as a whole). The friend group also showed a sign
days before the peak in daily incidence in the population
additional time to react to epidemics in small or large popula
on features of the outbreak and the network at hand. The
psychological, informational, or behavioral contagions that spread in networks.

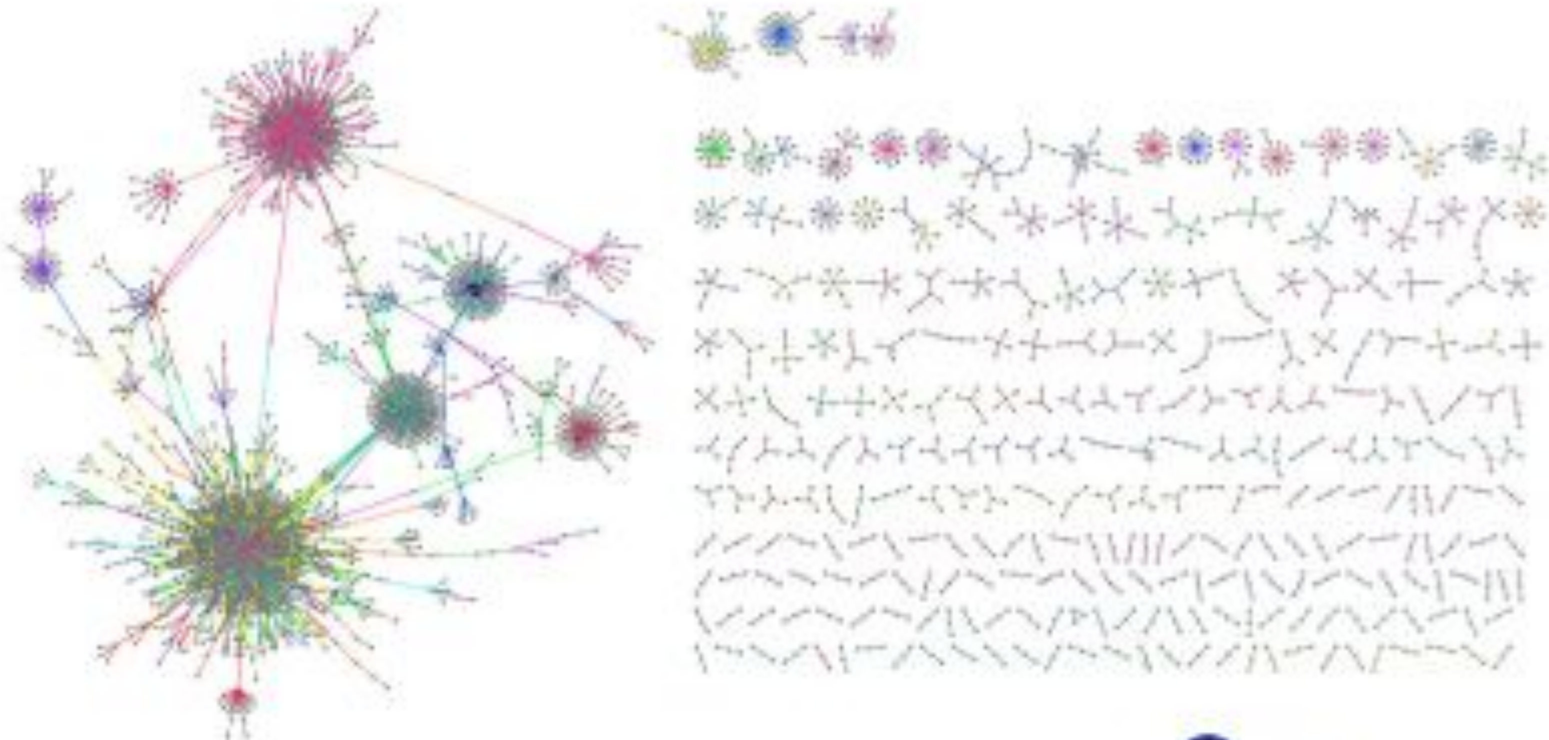
## Using **friendsensors**, can we detect **global contagious outbreaks** in **Twitter?**



476M tweets

2009 **Data** Kwak et al. (2010)

**~ 2/3 of Twitter**



Universidad Carlos III de Madrid @estebanmoro

Contagious outbreak?

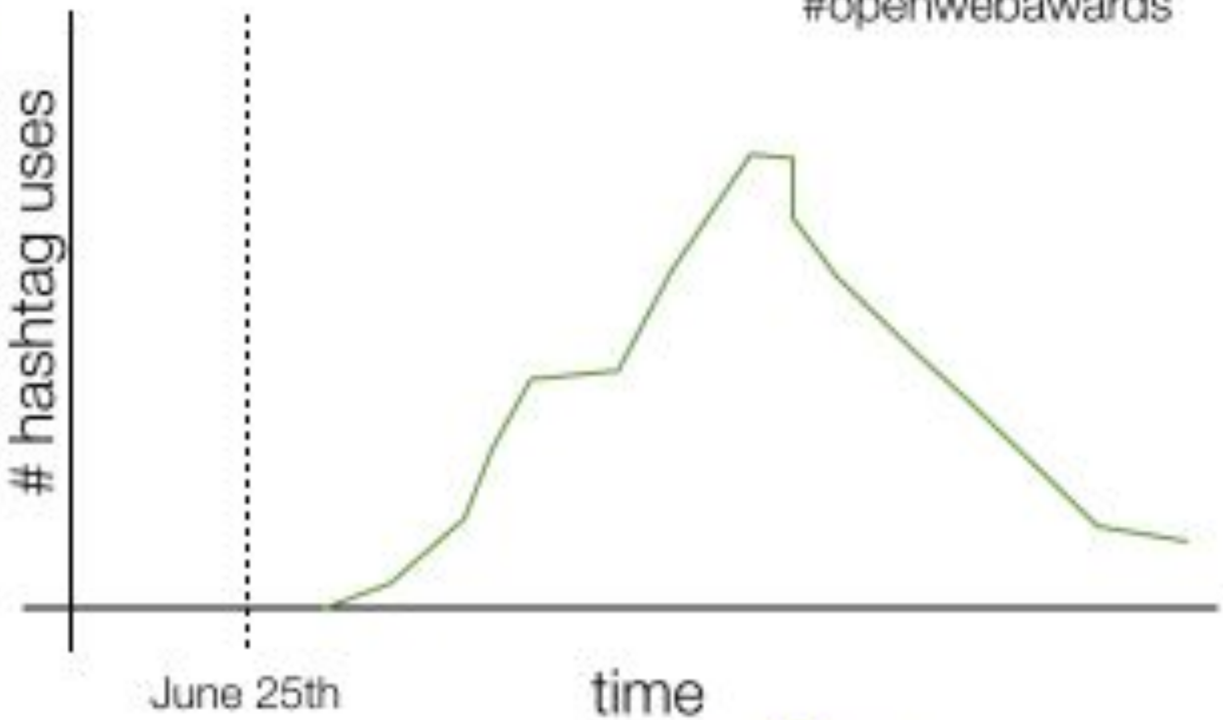
#indonesiaunite

#beatcancer #twitdraw

# Flu ≈ #hashtag

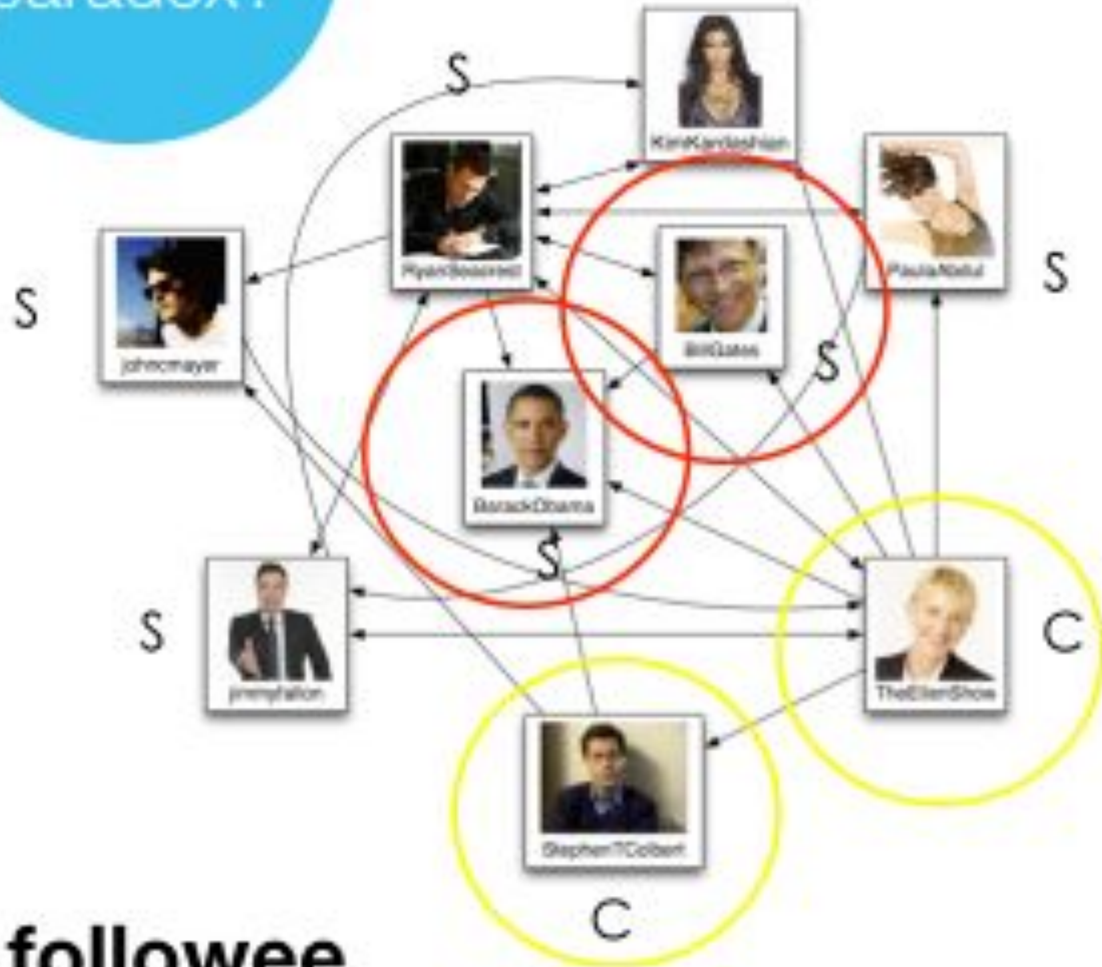
#pengakuan #lightupnigeria

#openwebawards




@estebanmoro

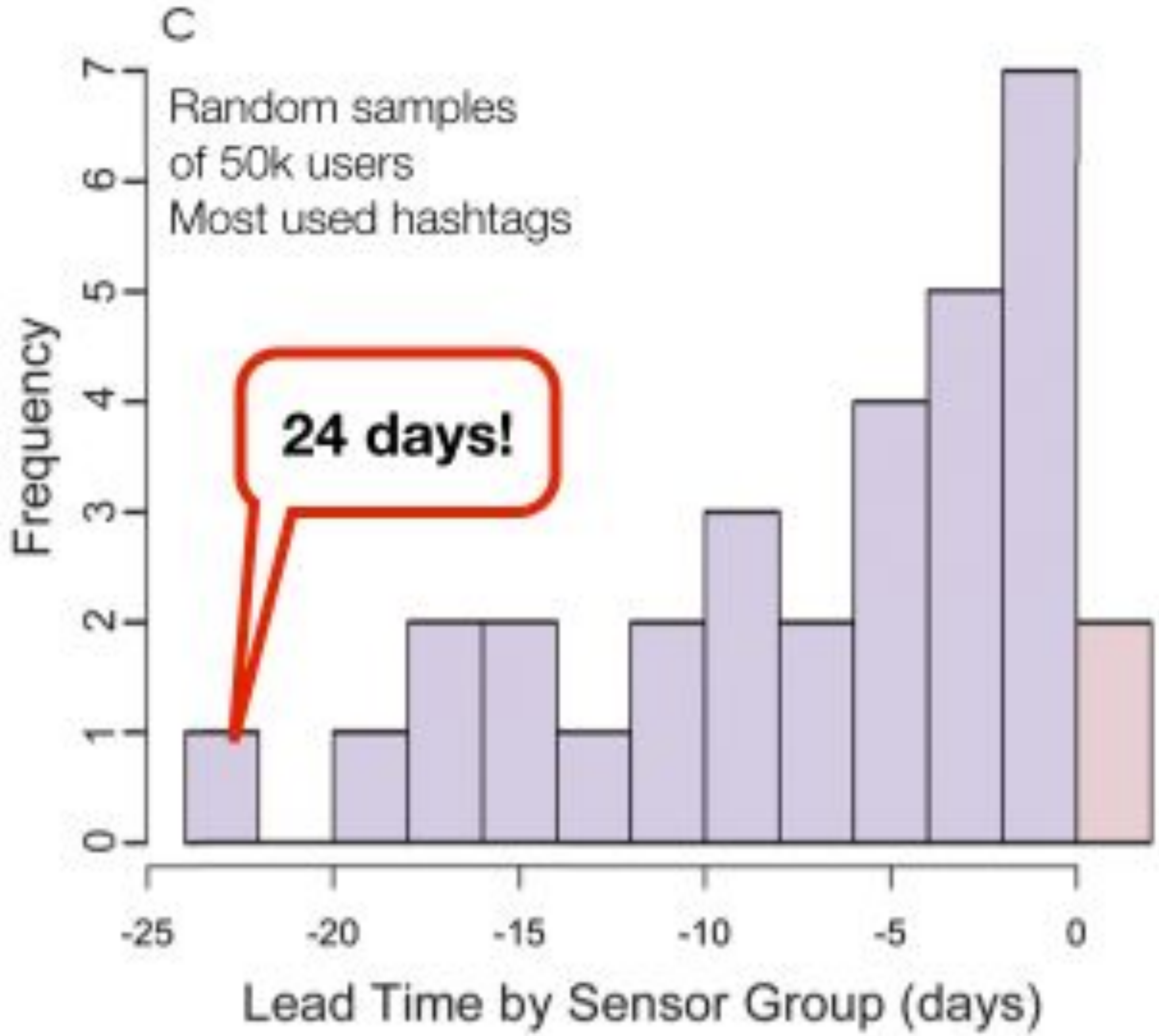
Friendship paradox?




**Friend = Follower / followee**

# Results: **Global view** (ex post)

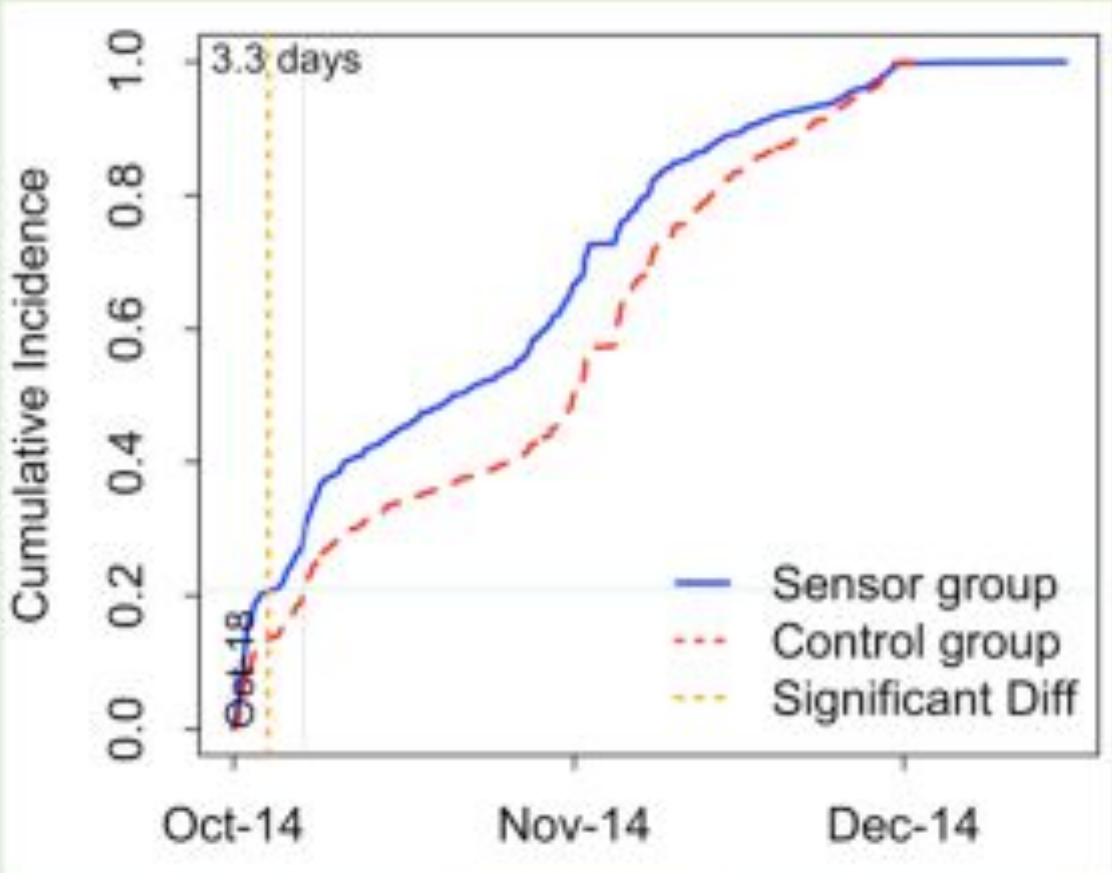
$$\langle t_{inf,i} \rangle_{i \in \text{Sensors}} \leq \langle t_{inf,i} \rangle_{i \in \text{Control}}$$



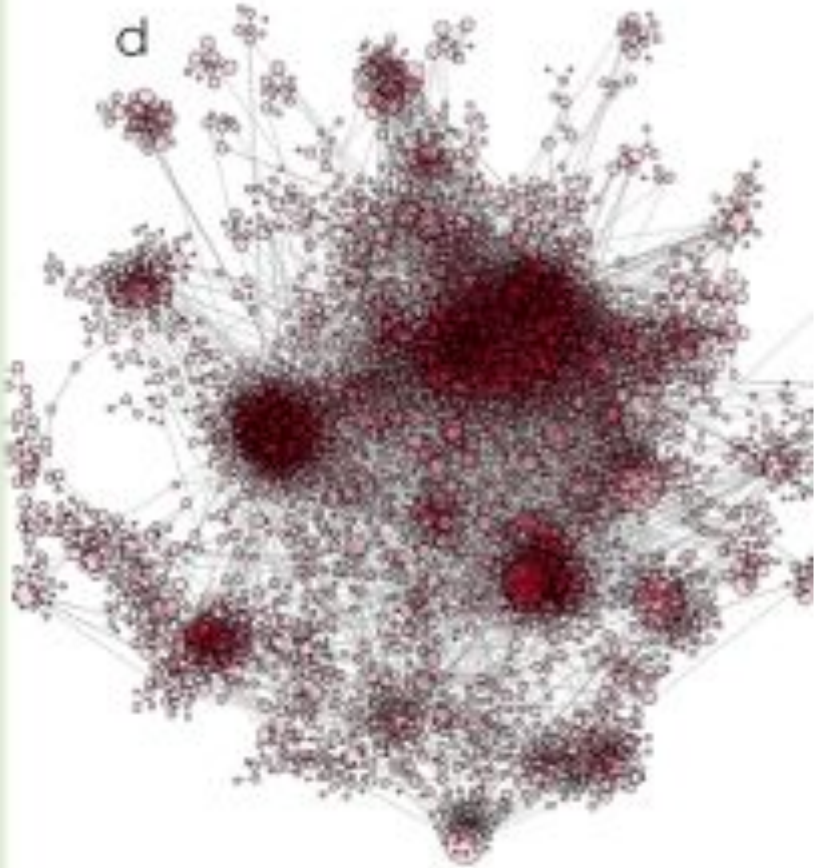

@estebanmoro

# Results: **early alarms** (real time)

- Divergence between cumulative incidence curves = early detecting of outbreaks

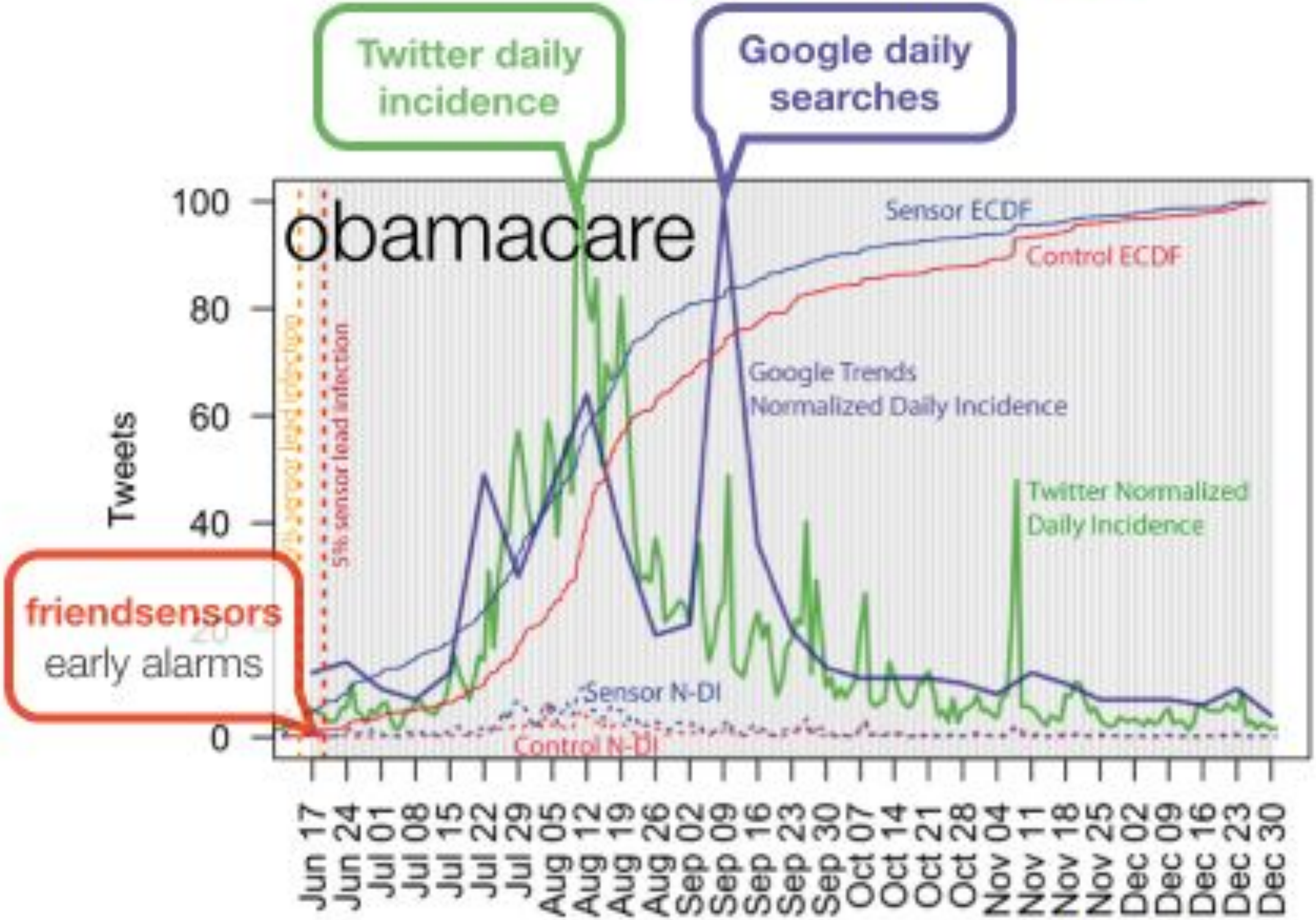



Twitter data
#openwebawards hashtag use



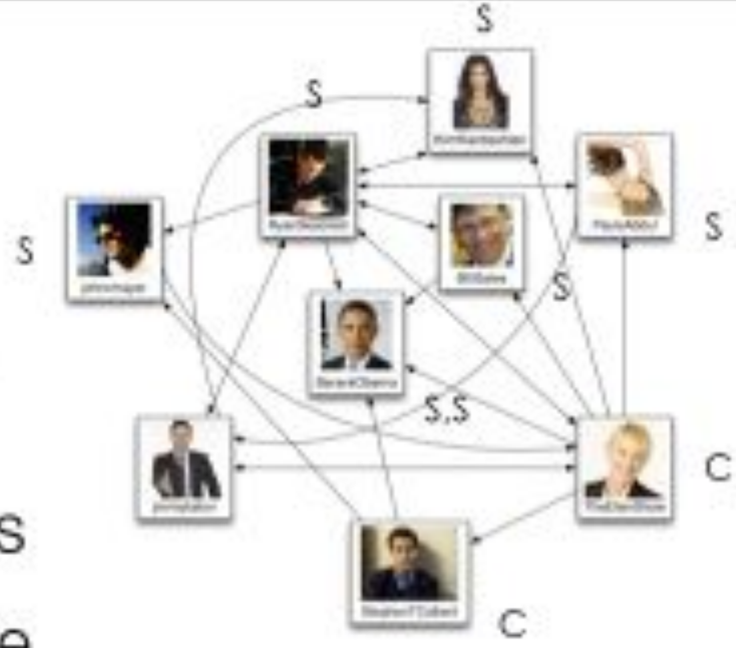

Universidad Carlos III de Madrid @estebanmoro

# Results: **but how well friendsensors work?**




@estebanmoro

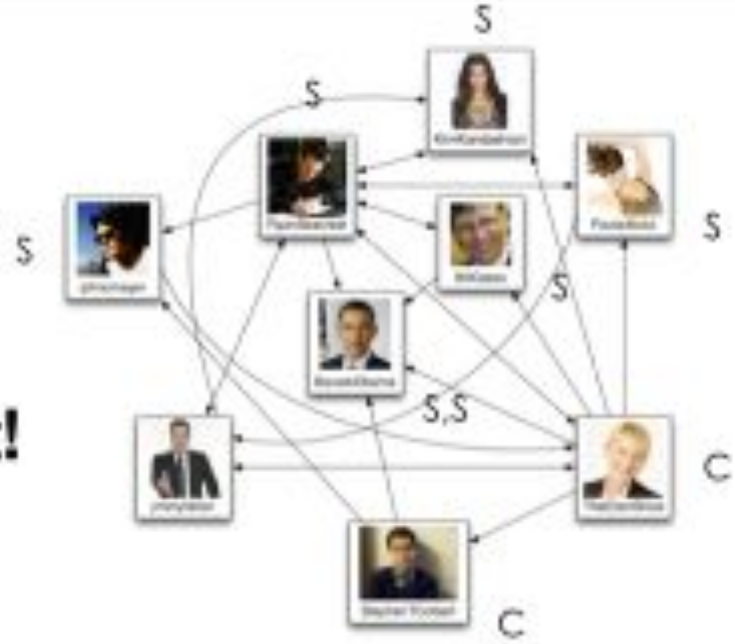
# Conclusions




- Monitoring social **BigData** requires a different approach
  - Local analysis for global conclusions
  - Take advantage of network structure
- Our **friendsensor** method **works** on Twitter
- Difference between control and sensors can be use to **early detection**

bit.ly/friendsensors

# Conclusions



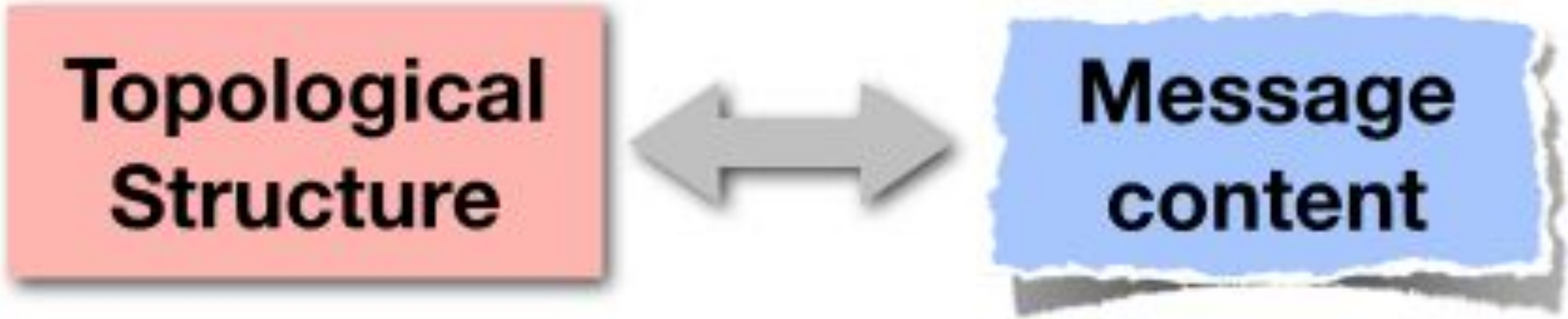

- Our **friendsensor** method has:
  - **Plenty of room for improvement!** We just used the simplest way to choose sensors.
  - Can be used in **other social networks**: is based on their social structure
  - Can be used for **highly dynamical scenarios**: choose sensors dynamically.
  - Can be adapted for **geographical filters, languages, interests**, etc: choose sensors accordingly.

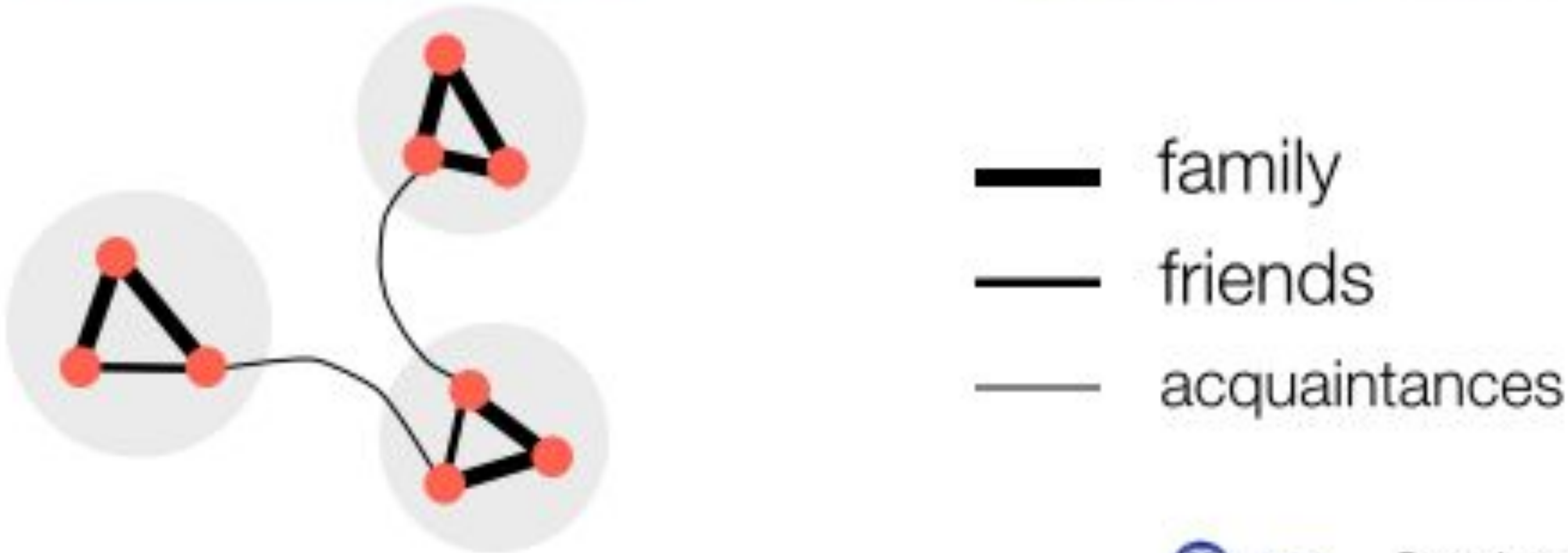
bit.ly/friendsensors

# What and where do we share on social networks? 4

# Motivation

- Grannoveter's theory '73: ***the strength of weak ties***



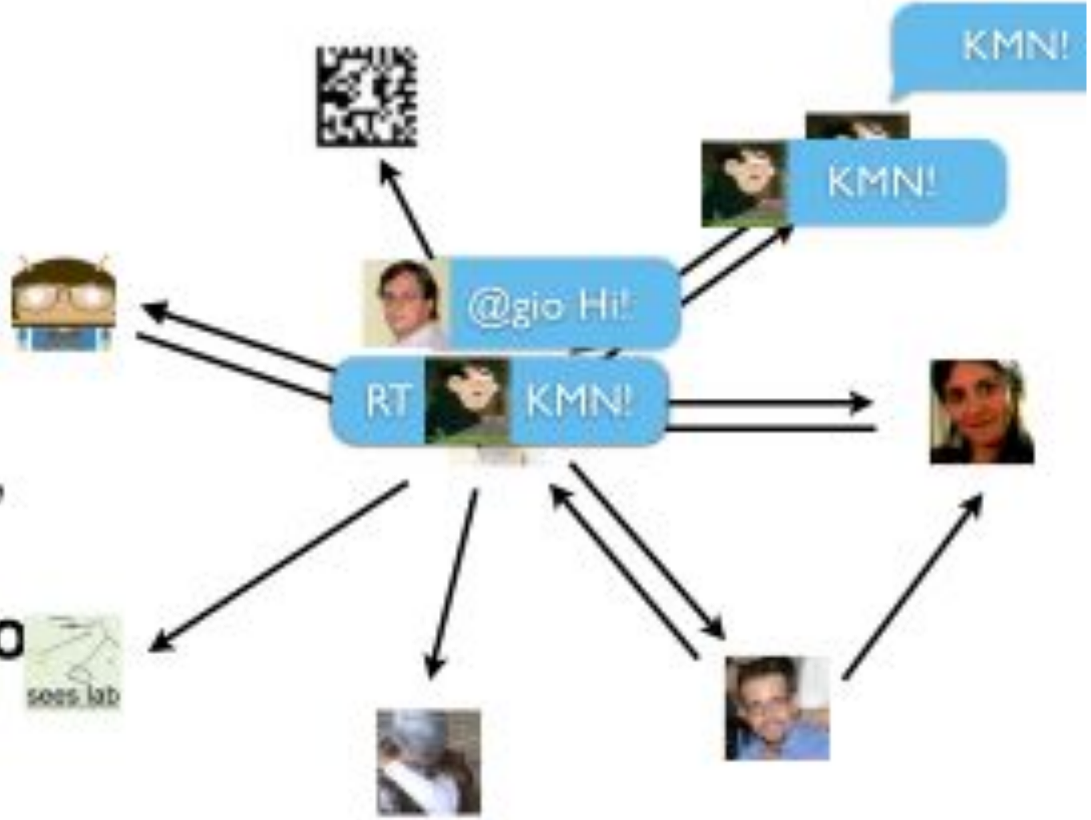



# Motivation

- Can we test these ideas in online social networks?
- Online Social networks:
  - Data availability
  - Large scale
  - Different types of interactions are explicitly separated
    - information diffusion events
    - personal communications
  - Easy interaction

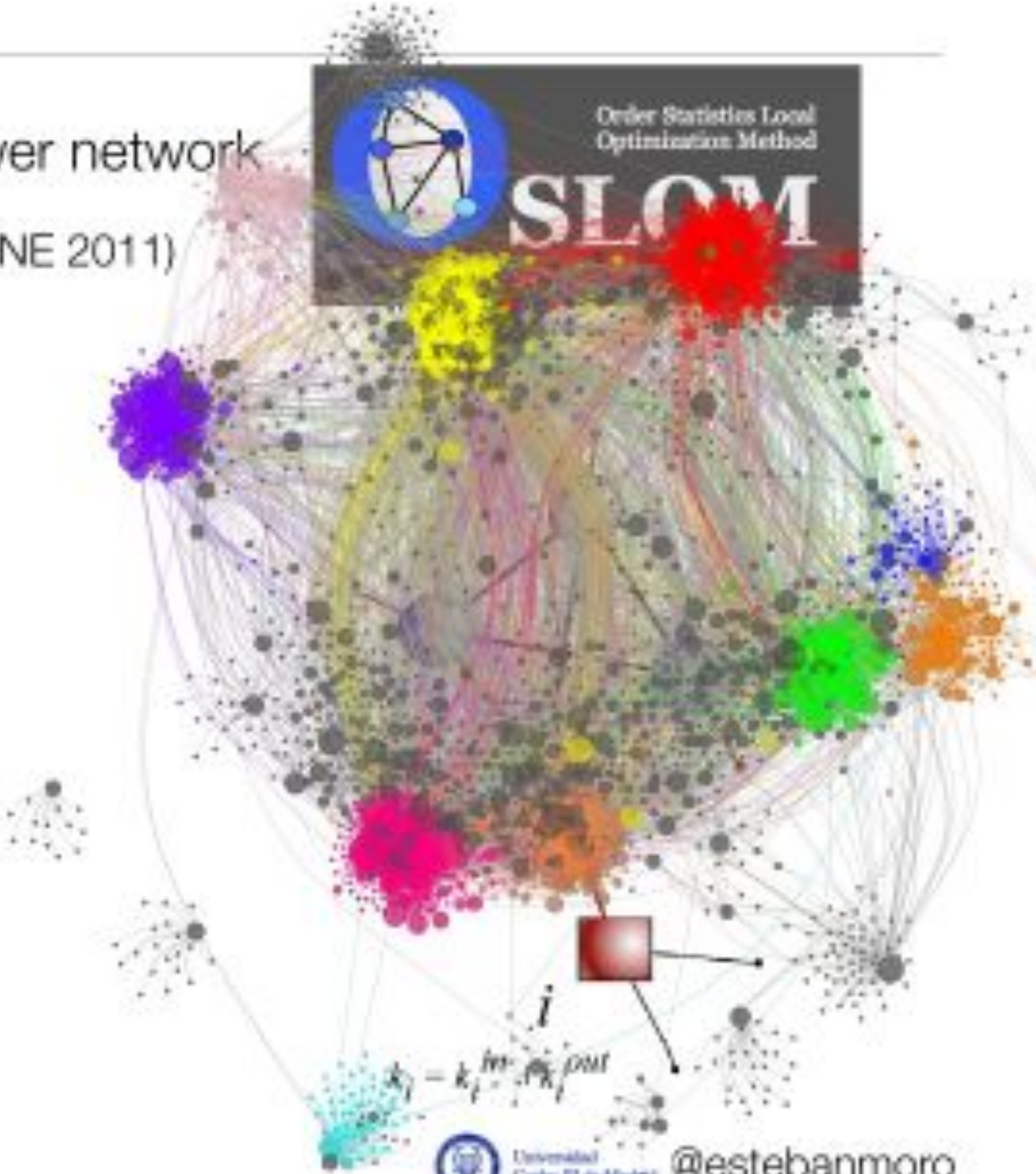
# Motivation

- Twitter
  - **Relationship**:
    - Followers / following
  - **Interaction:**
    - Mentions, replies: “@”
  - **Diffusion of informatio**
    - Retweets: “RT”



# Groups

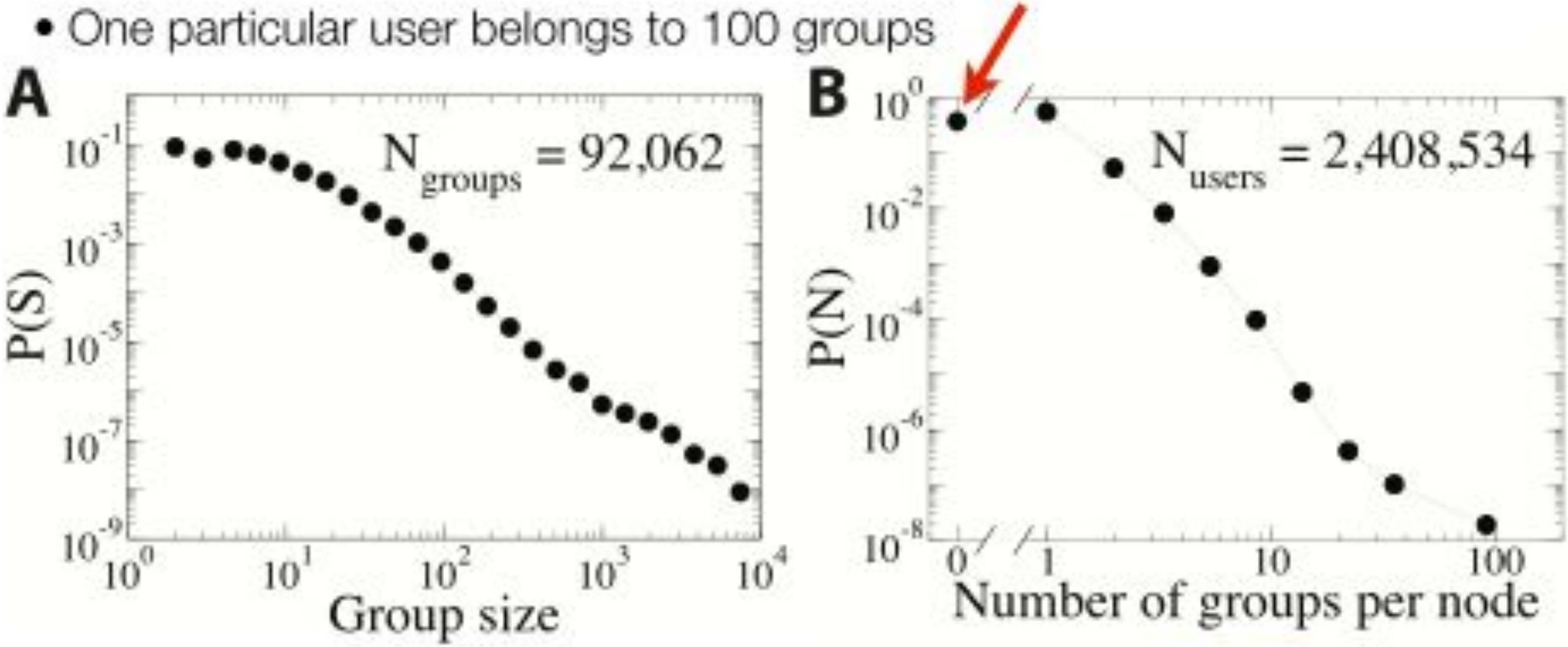
- Community analysis on the follower network
  - OSLOM (Lancichinetti et al, PLoS ONE 2011)
    - Large network
    - Directed network
    - Users in multiple groups
  - Other methods used:
    - Infomap
    - Moses
    - Louvain



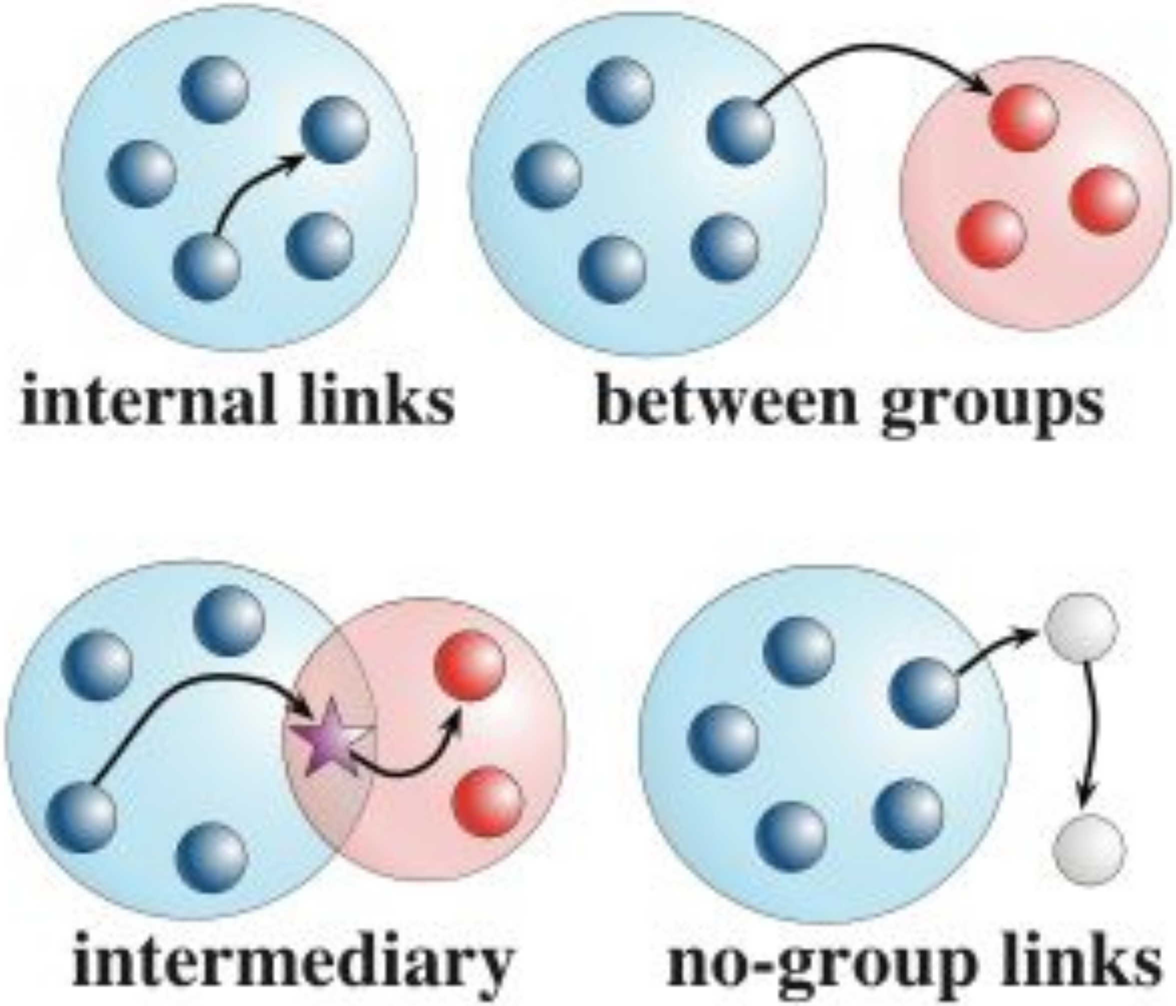
# Groups

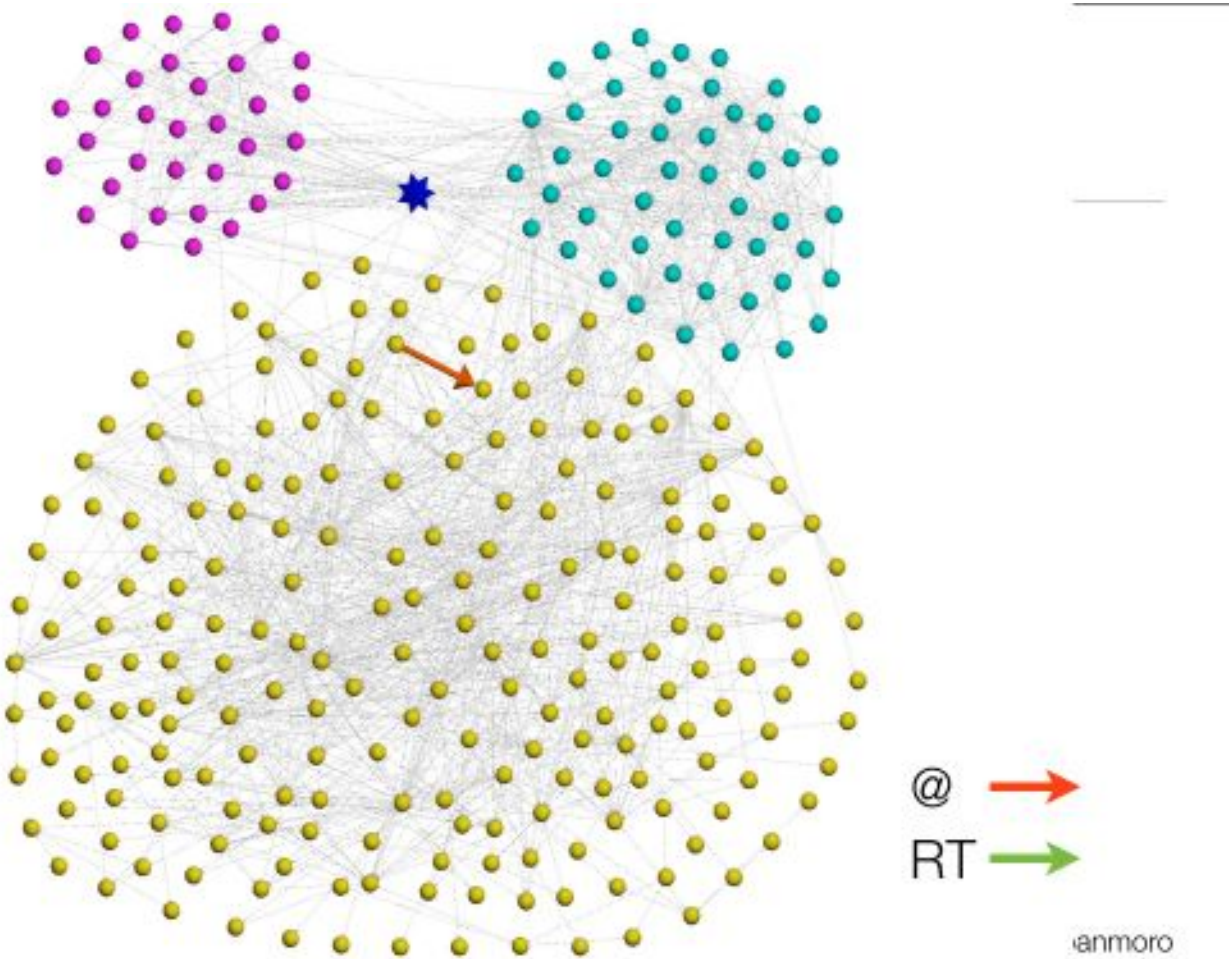
Grabowick , Ramasco, Moro, 20121

- 92062 Groups
- Nodes can be "homeless"
  - 37.4% of accounts not allocated to any group
- One particular user belongs to 100 groups



# Links/interactions between/inside groups



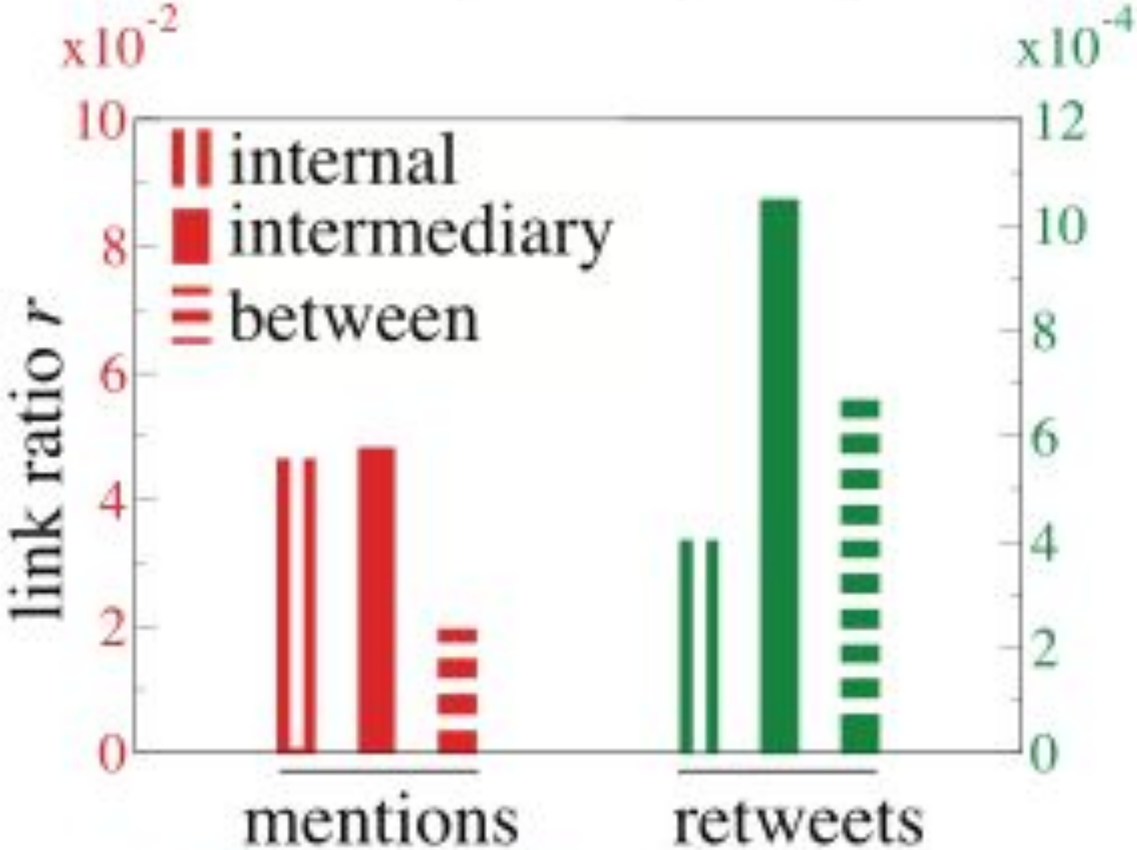

@
RT
anmoro

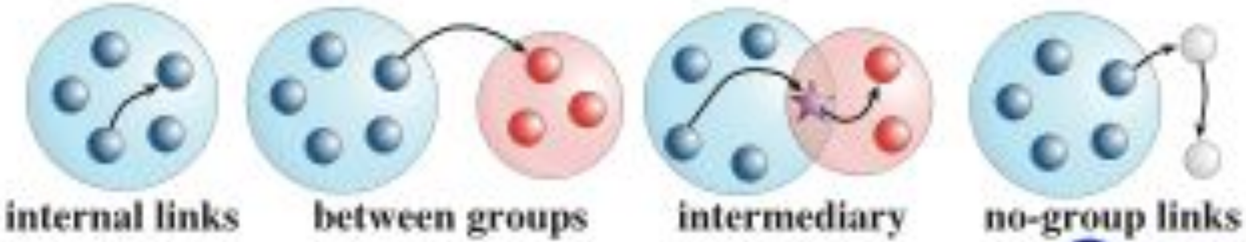
# Links between groups

Grabowick , Ramasco, Moro, 20121

- Strong **Interaction (@)** happens mostly inside groups

- **Diffusion information (RT)** happens mostly at intermediary links



# Discussion

| Offline social networks | Twitter |
|---|---|
| Strong social ties tend to appear at the interior of groups | Stronger mentions are more likely to happen on internal links |
| Weak ties are expected to be more common connecting different groups | Mentions are less likely to happen on intermediary and between groups |
| Links between groups are more important for information diffusion | Retweets happen mostly between groups |

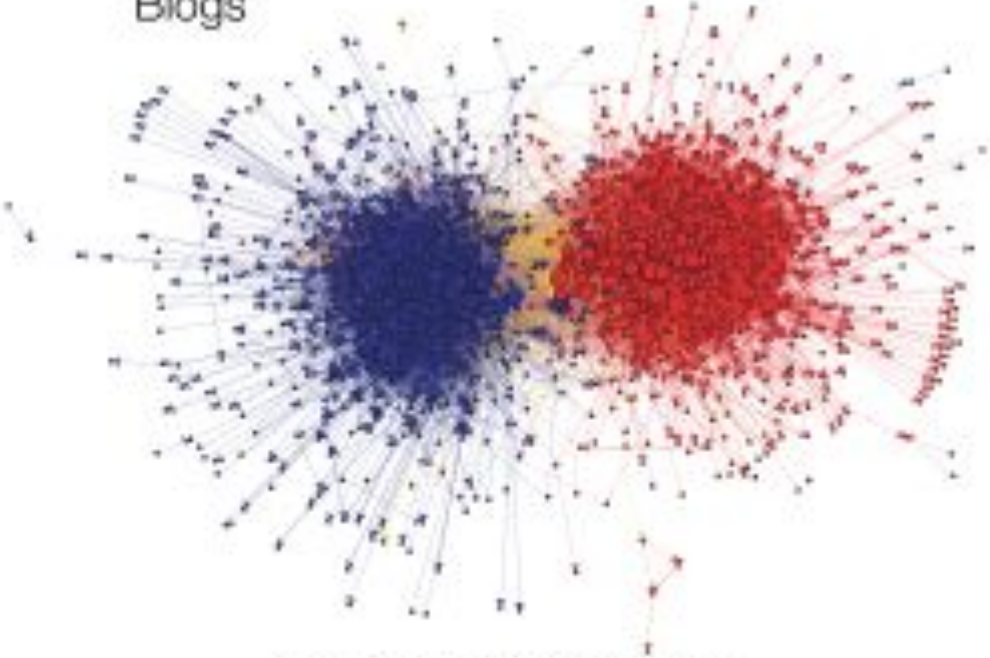
# 5

# How does political opinion propagates in social networks?

# Political conversation in social media

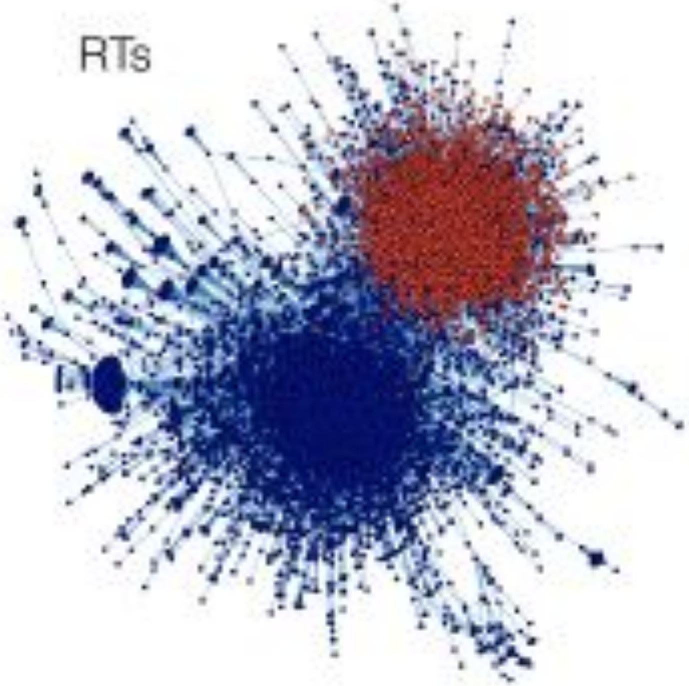
## High polarization of the conversation

Blogs



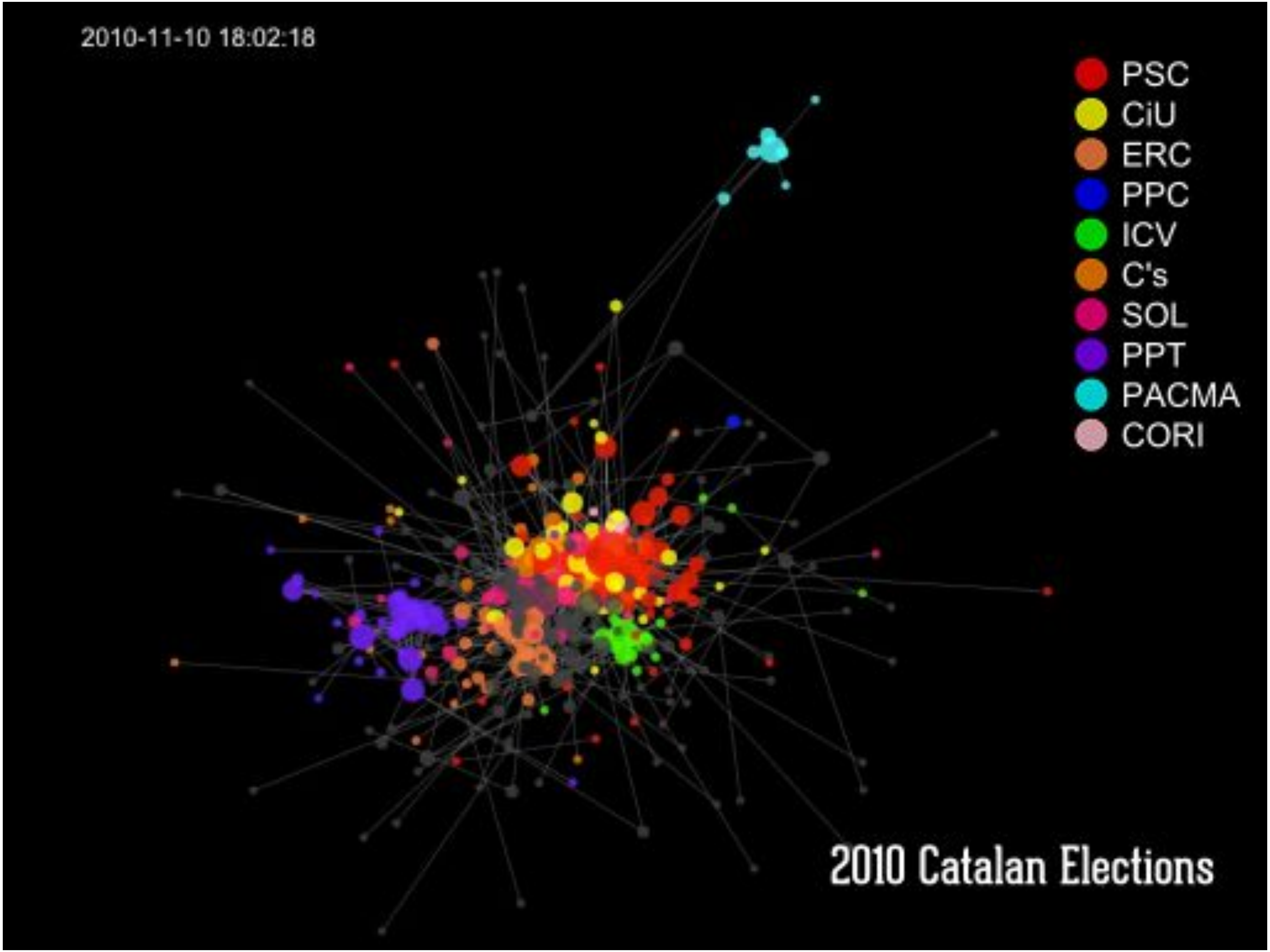
Adamic & Glance 2005

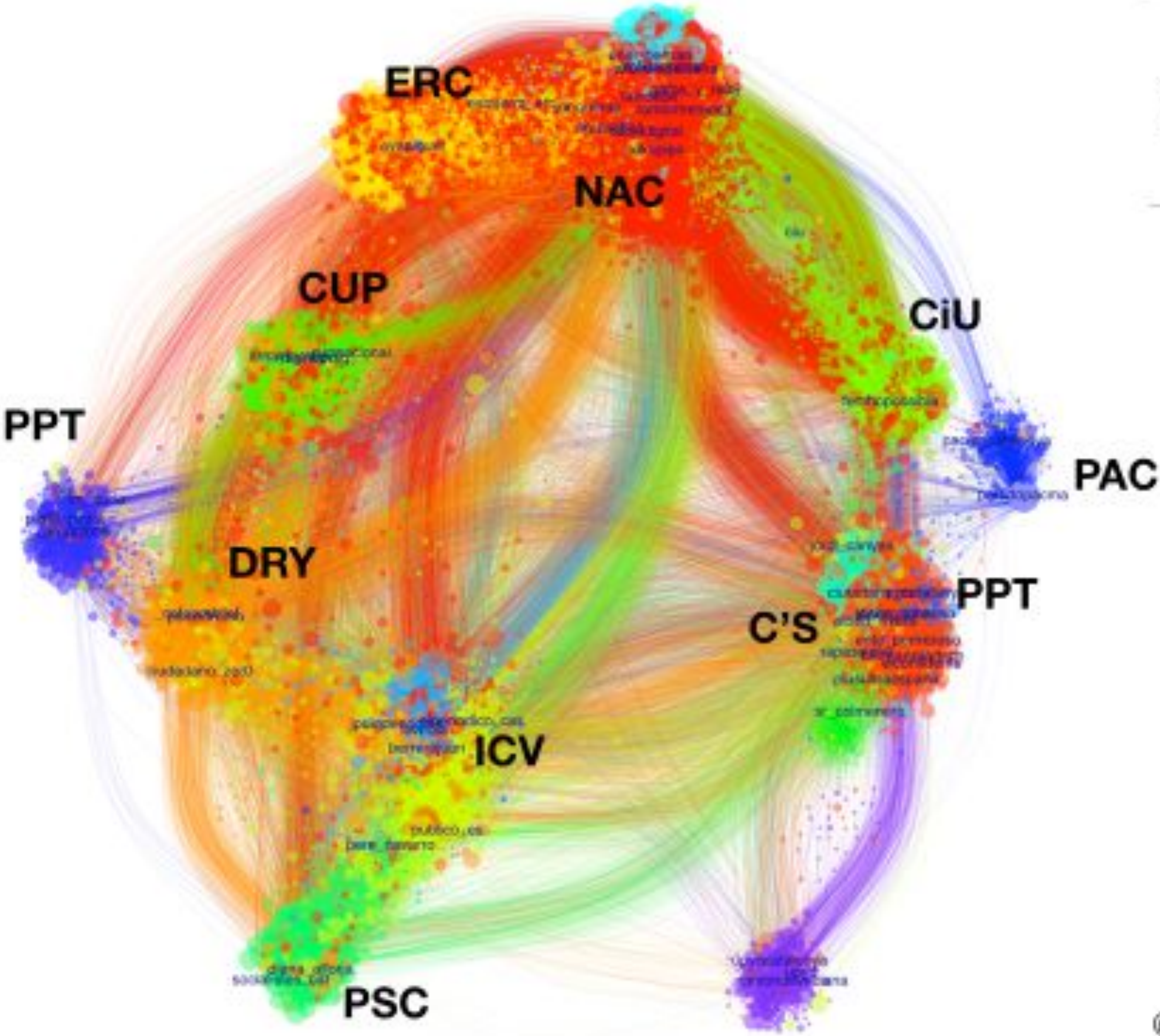
RTs



Conover et al 2012

@estebanmoro


2010-11-10 18:02:18
PSC
CiU
ERC
PPC
ICV
C's
SOL
PPT
PACMA
CORI
2010 Catalan Elections

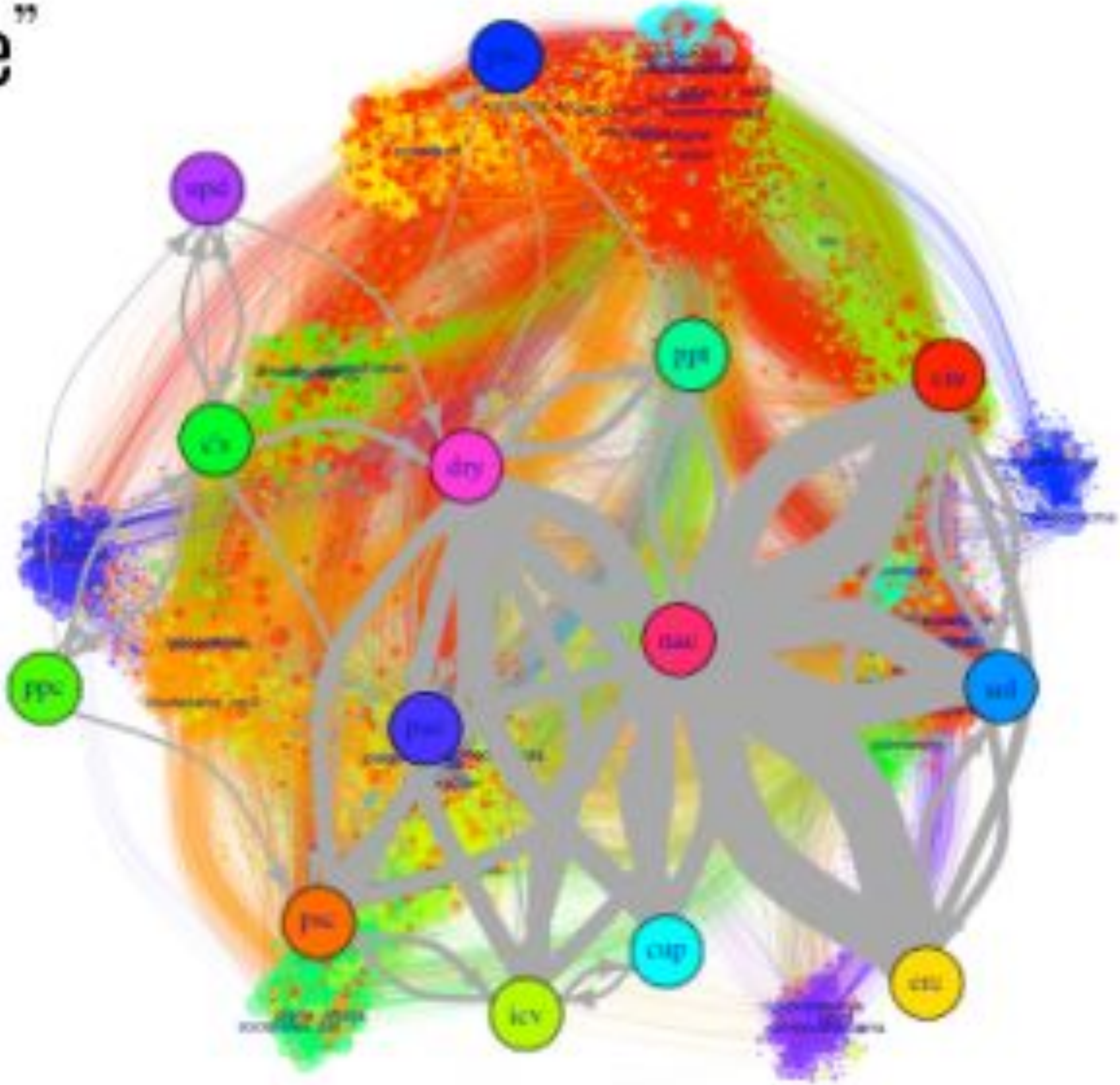



2012

@estebanmoro

# Communication between communities

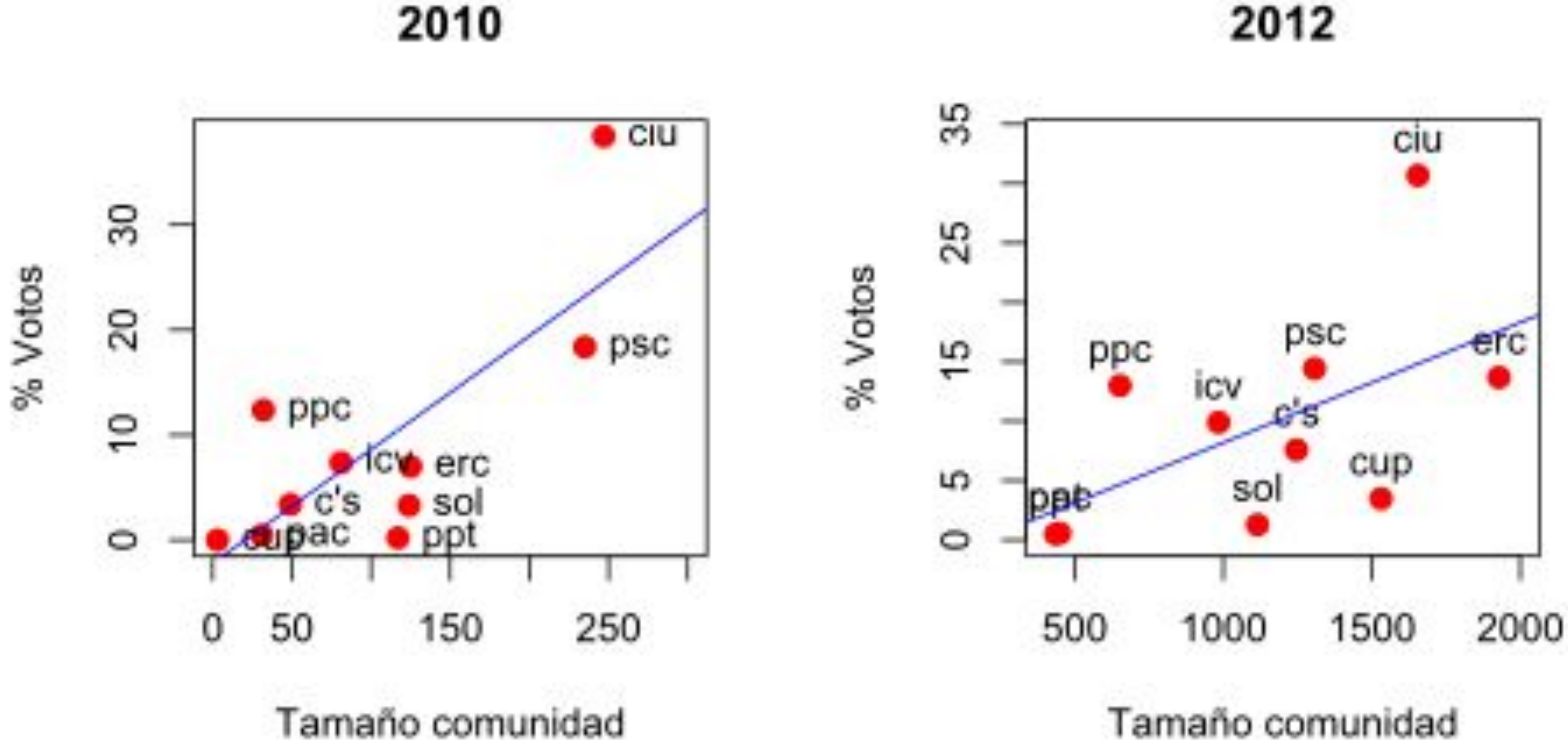
## "Ideological space"



@estebanmoro

# Communities and votes

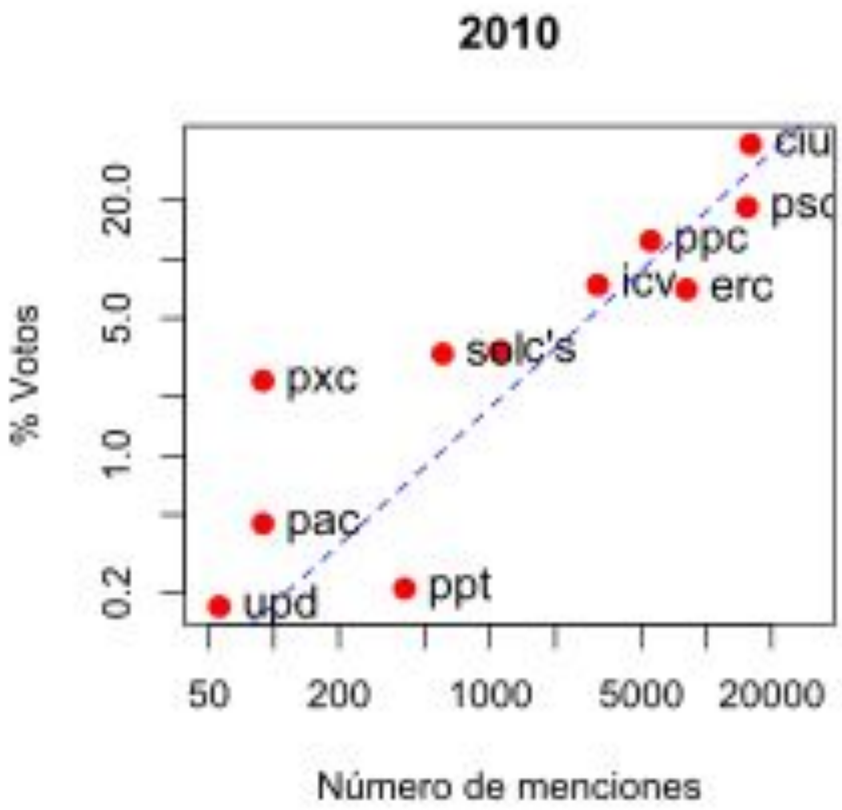
Is there a relationship between community size and party votes?

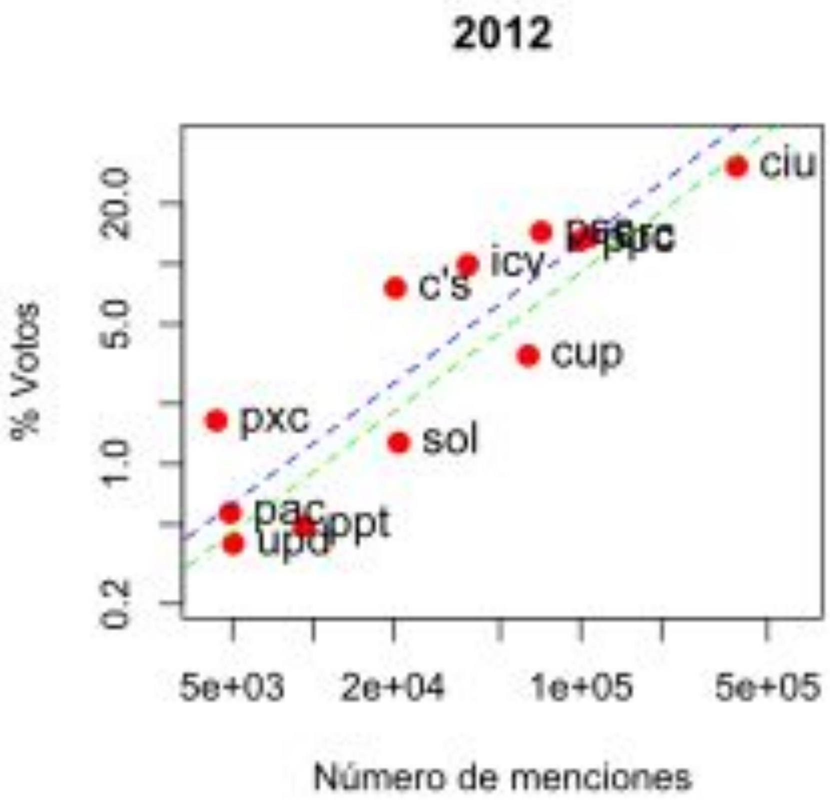



@estebanmoro

# Mentions and votes

## Is there a relationship between mentions and party votes?

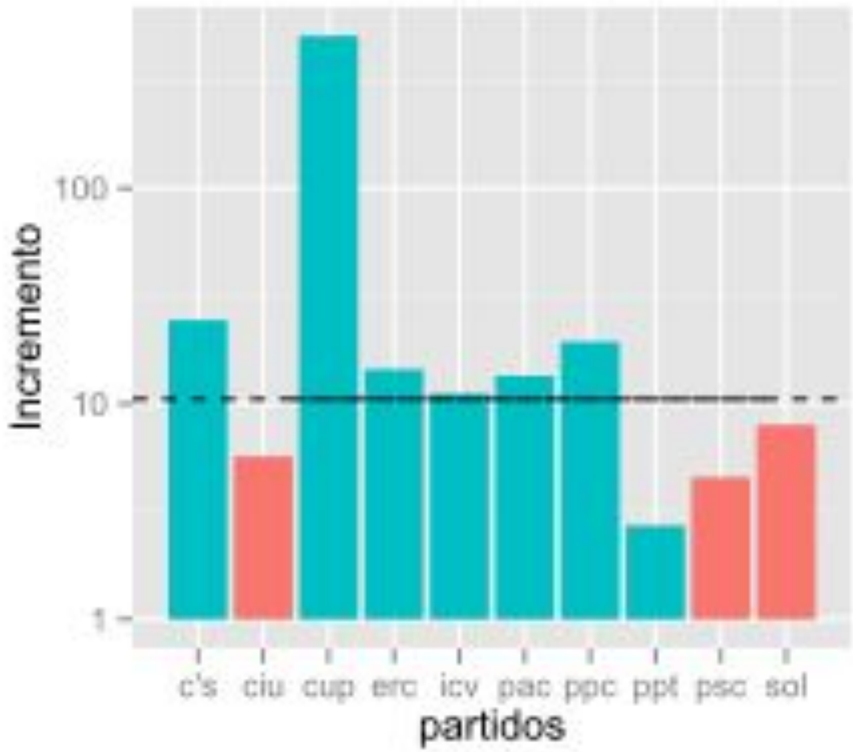


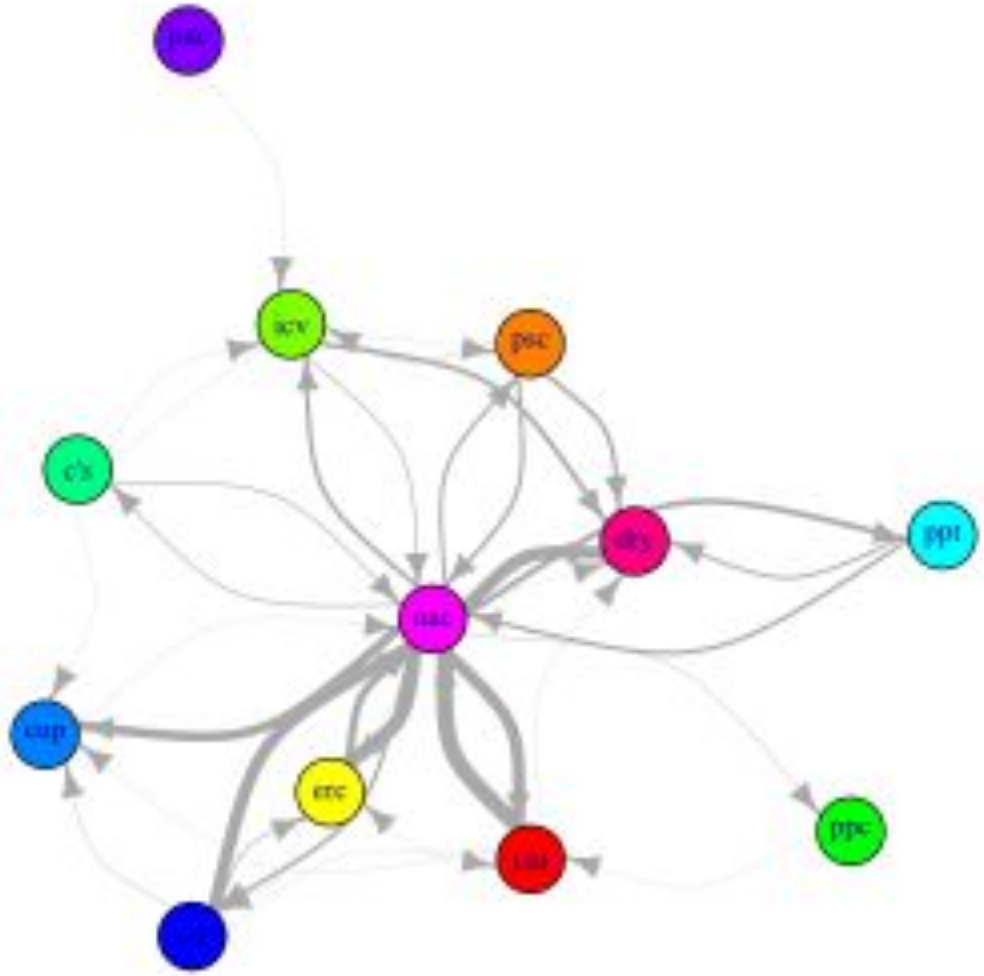




@estebanmoro

# Communities evolve

## They change in [relative] size and users move between them



# Summary

Congosto & Moro 2011/13

1. Polarization of political communication: one party, one community

2. Communication between ideologically close communities only

3. Communities size =/= votes

4. Communities are very persistent. Exchange between communities are only between those ideologically close.


@estebanmoro