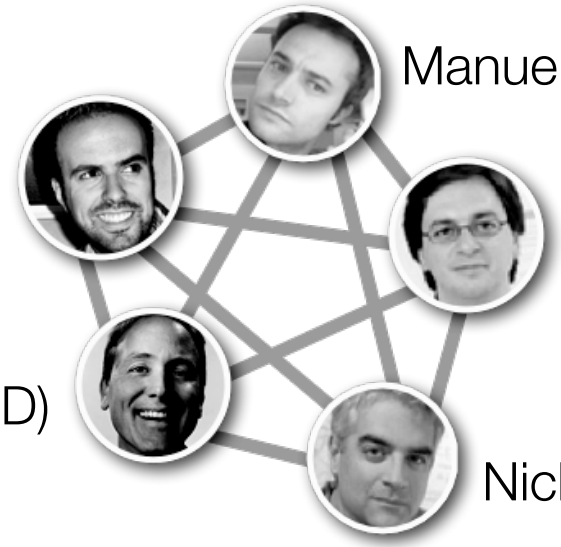
Using

friends as sensors

to **detect** global-scale contagious outbreaks



Manuel García-Herranz (UAM)

Manuel Cebrián (NICTA)

Esteban Moro (UC3M & IIC)

James Fowler (UCSD)

Nicholas Christakis (Yale)

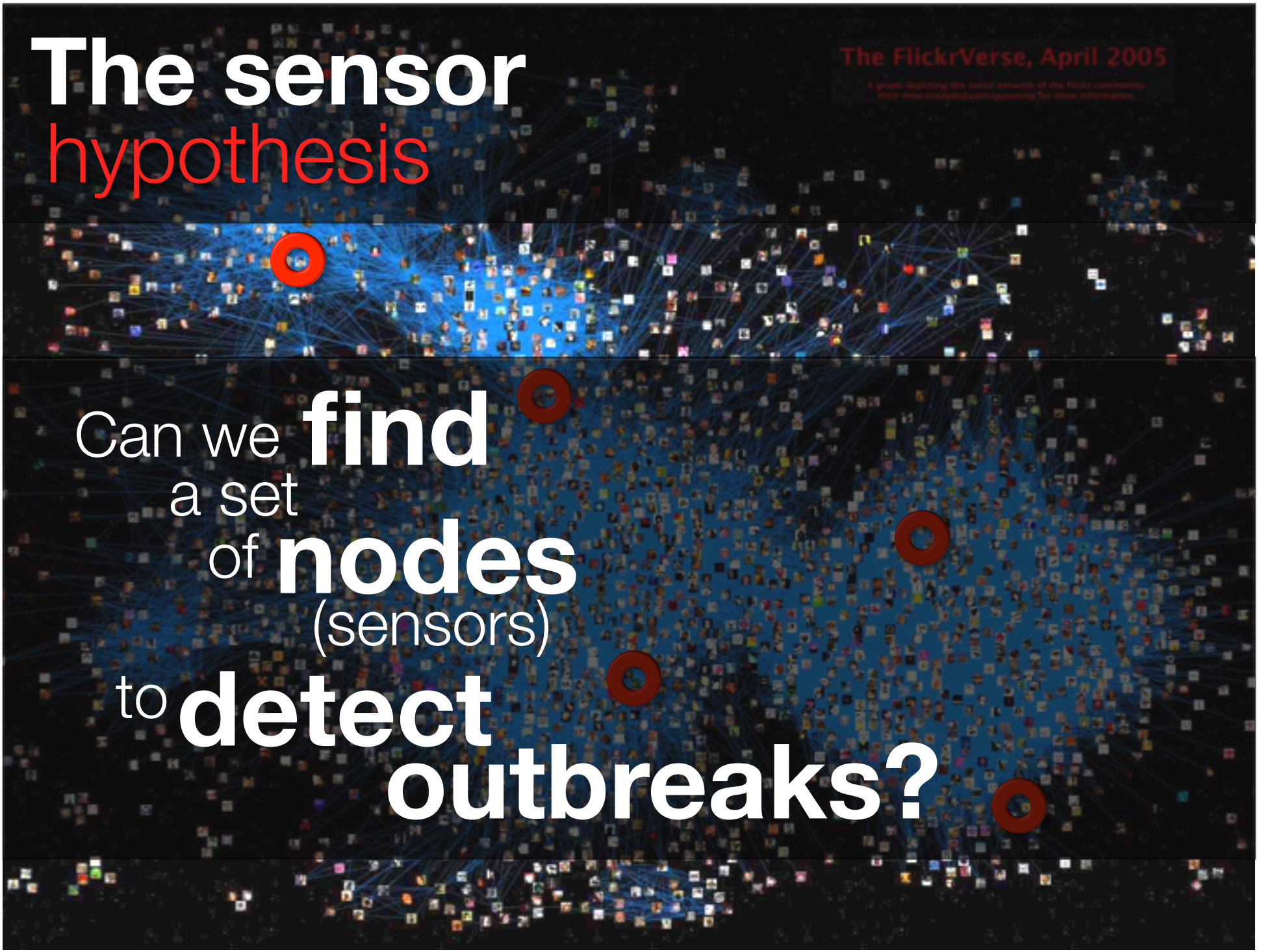
bit.ly/friendsensors

1 outbreak detection

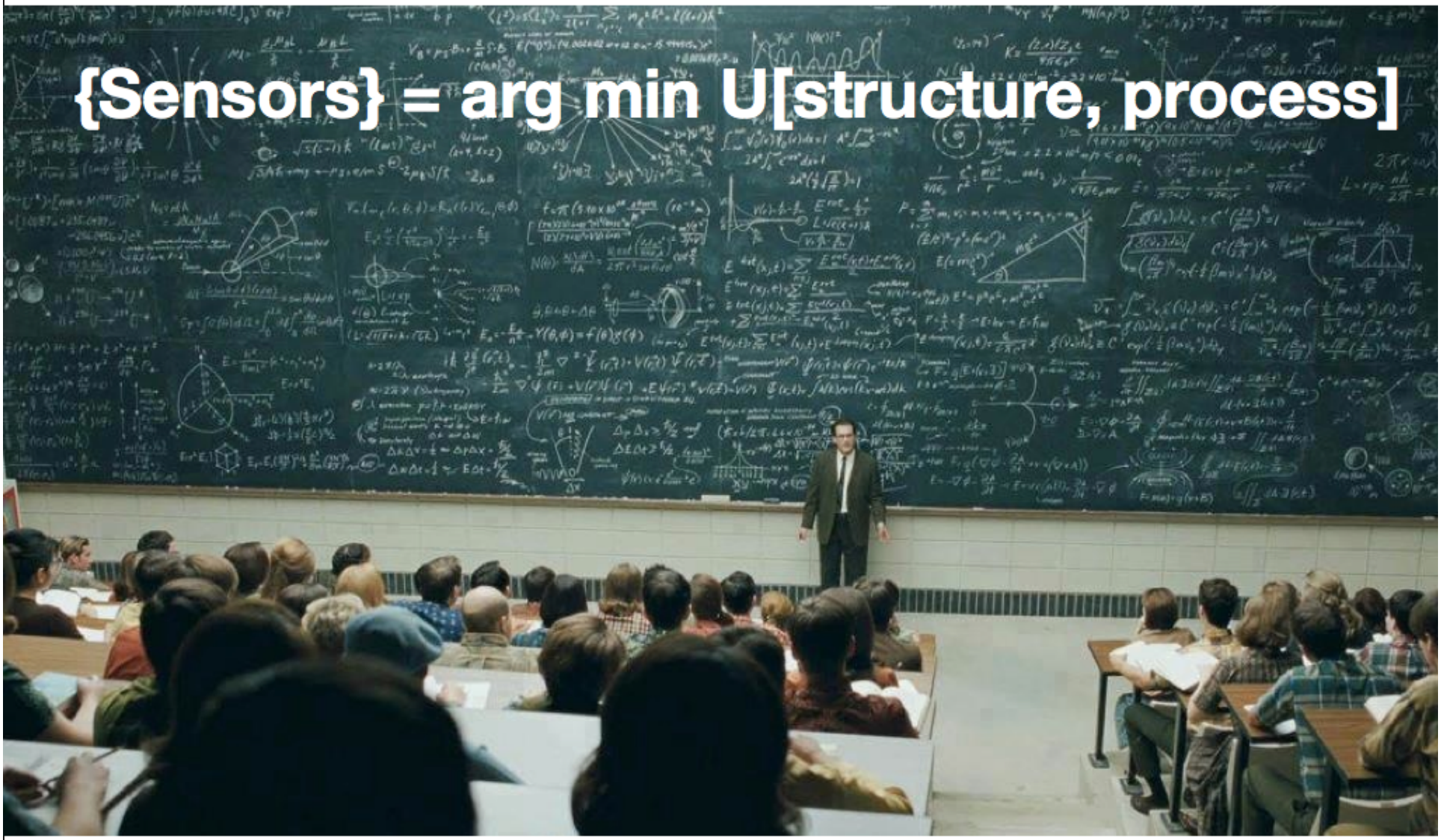
in networks

Outbreak detection
in networks

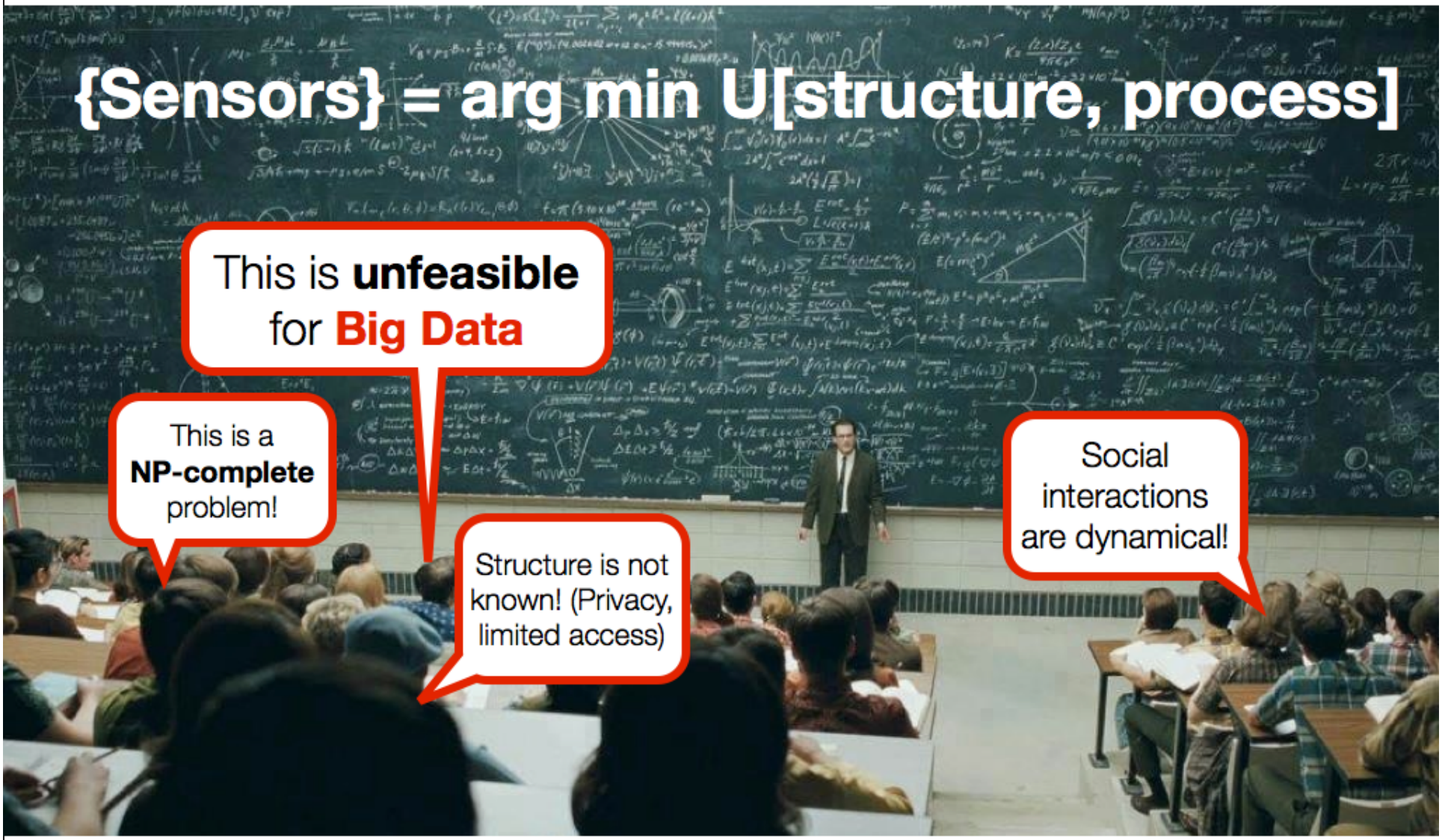
Can we monitor the whole network?


The FlickrVerse, April 2005
The sensor
hypothesis
Can we find
a set
of nodes
(sensors)
to detect
outbreaks?

Sensor hypothesis



Sensor hypothesis

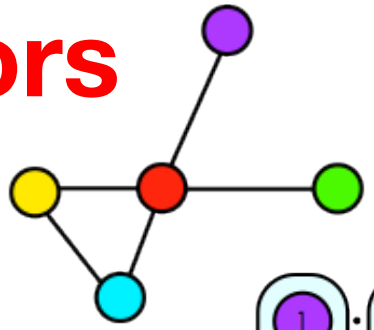


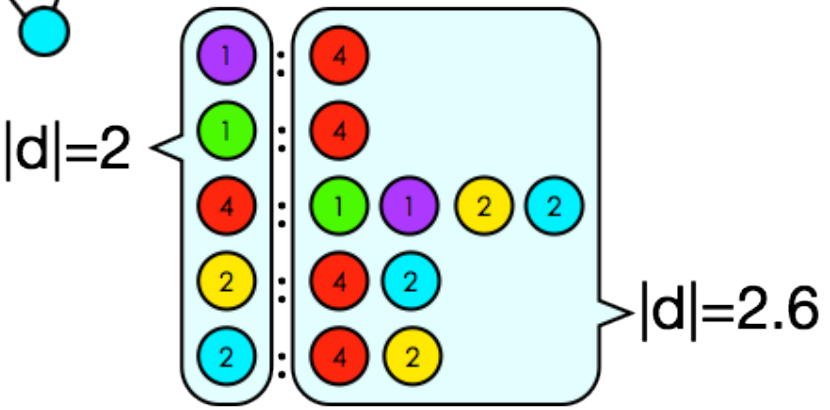
A solution:
Friends as Sensors

Use the friendship paradox: the friends of a set of users are more central

A solution:

Friends as Sensors





Why Your Friends Have More Friends than You Do[1]

Scott L. Feld
State University of New York at Stony Brook

It is reasonable to suppose that individuals use the number of friends that their friends have as one basis for determining whether they, themselves, have an adequate number of friends. This article shows that, if individuals compare themselves with their friends, it is likely that most of them will feel relatively inadequate. Data on friendship drawn from James Coleman's (1961) classic study *The Adolescent Society* are used to illustrate the phenomenon that most people have fewer friends than their friends have. The logic underlying the phenomenon is mathematically explored, showing that the mean number of friends of friends is always greater than the mean number of friends of individuals. Further analysis shows that the proportion of individuals who have fewer friends than the mean number of friends their own friends have is affected by the exact arrangement of friendships in a social network. This disproportionate experiencing of friends with many friends is related to a set of abstractly similar "class size paradoxes" that includes such diverse phenomena as the tendencies for college students to experience the mean class size as larger than it actually is and for people to experience beaches and parks as more crowded than they usually are.

Friendship is not only a source of satisfaction and security; it is also a way that individuals evaluate themselves and others. People expect themselves and others to have friends and wonder about the normality of those individuals who appear to have few or no friends. There has

You:

$$\mu = \frac{\sum_{v \in V} d(v)}{|V|} = \frac{2|E|}{|V|}.$$

Your friends:

$$\frac{\sum_{uv \in E} d(v)}{2|E|} = \mu + \frac{\sigma^2}{\mu},$$

A solution: **FriendSensors**

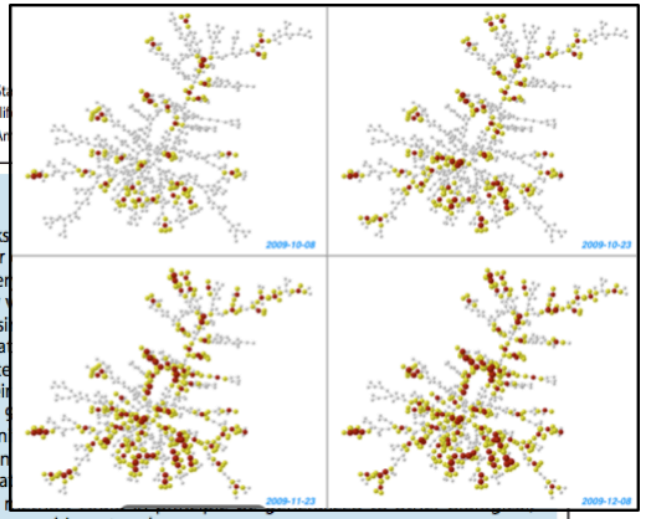
OPEN ACCESS Freely available online

PLoS one

Social Network Sensors for Early Detection of Contagious Outbreaks

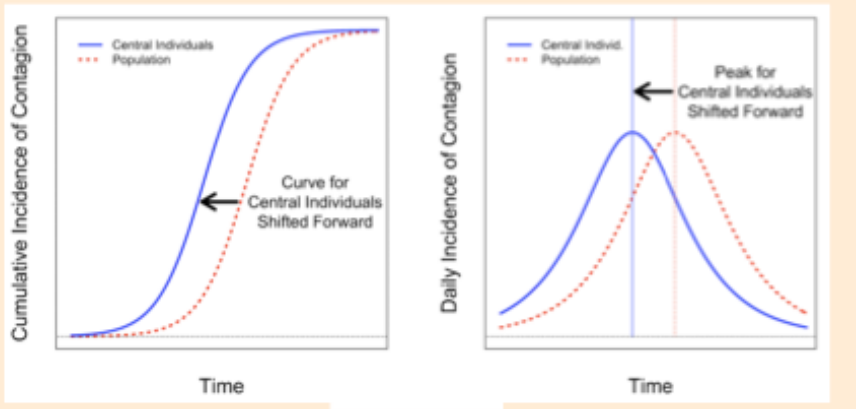
Nicholas A. Christakis[1,2]*, **James H. Fowler**[3,4]

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 Ar



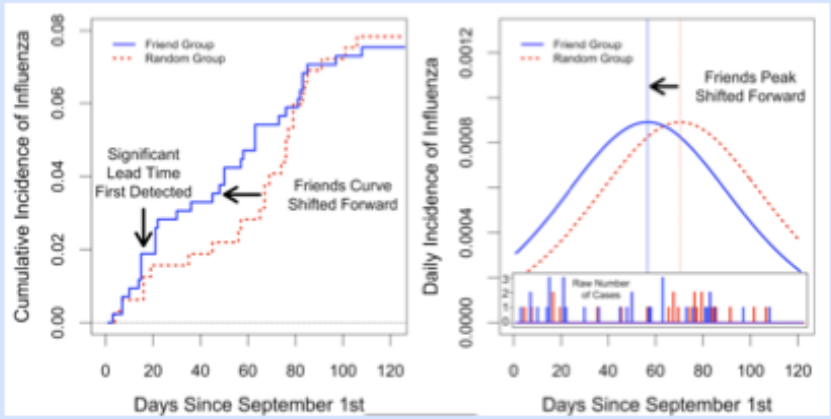
Abstract

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 peripher
individuals who might be monitored for infection is typically
require ascertainment of global network structure, namely, si
Such individuals are known to be more central. To evaluat
detection, we studied a flu outbreak at Harvard College in late
a group of randomly chosen individuals or a group of their
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.



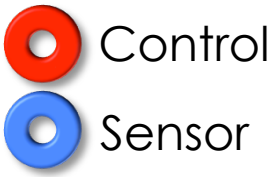
expected

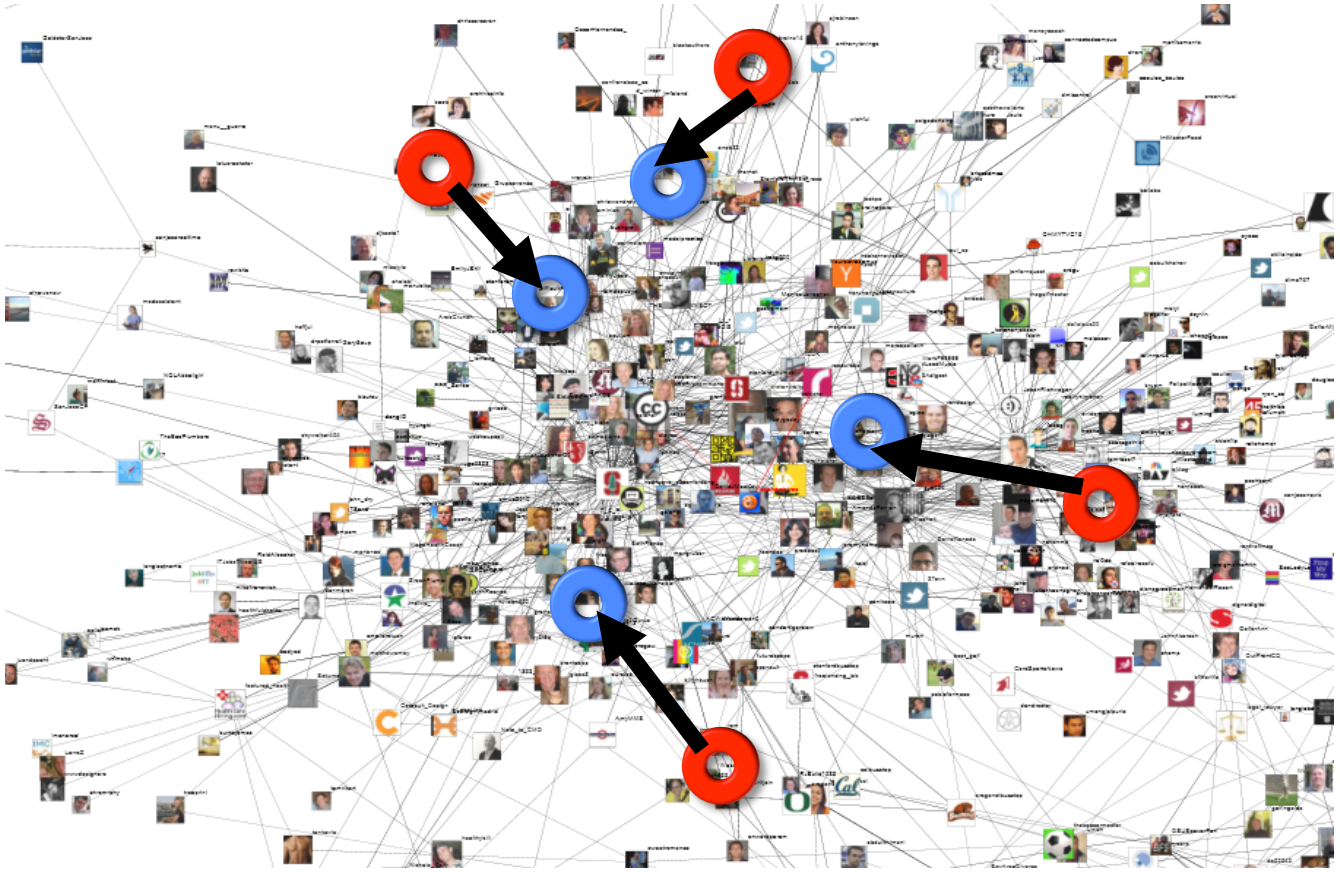
observed



A solution:

Friends as Sensors



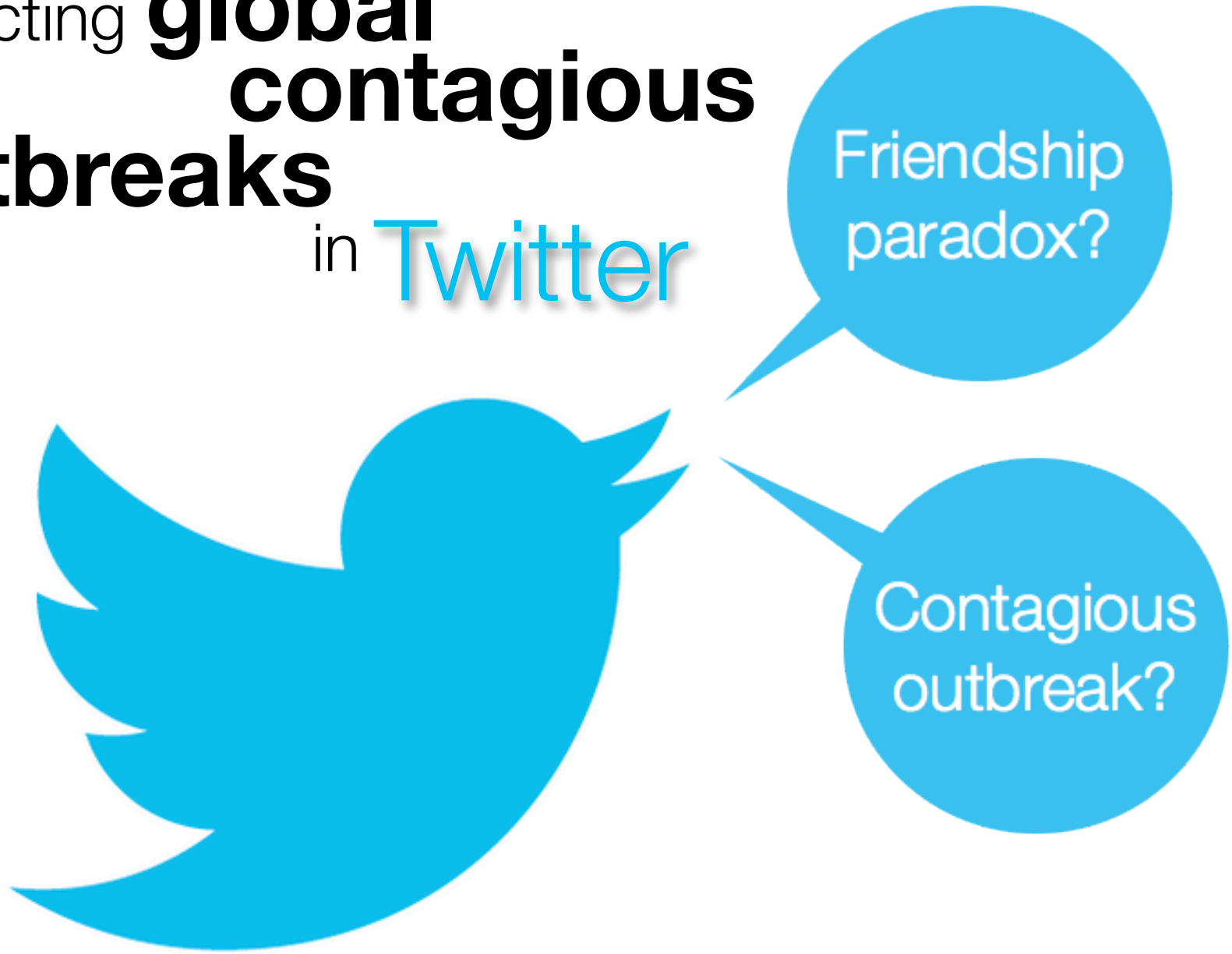


$$\langle t_{inf,i} \rangle_{i \in \text{Sensors}} \leq \langle t_{inf,i} \rangle_{i \in \text{Control}}$$

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

2 outbreak detection in online social networks

Detecting **global contagious outbreaks** in Twitter



476M tweets

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

~ 2/3 of Twitter

contagious outbreak?

#indonesiaunite

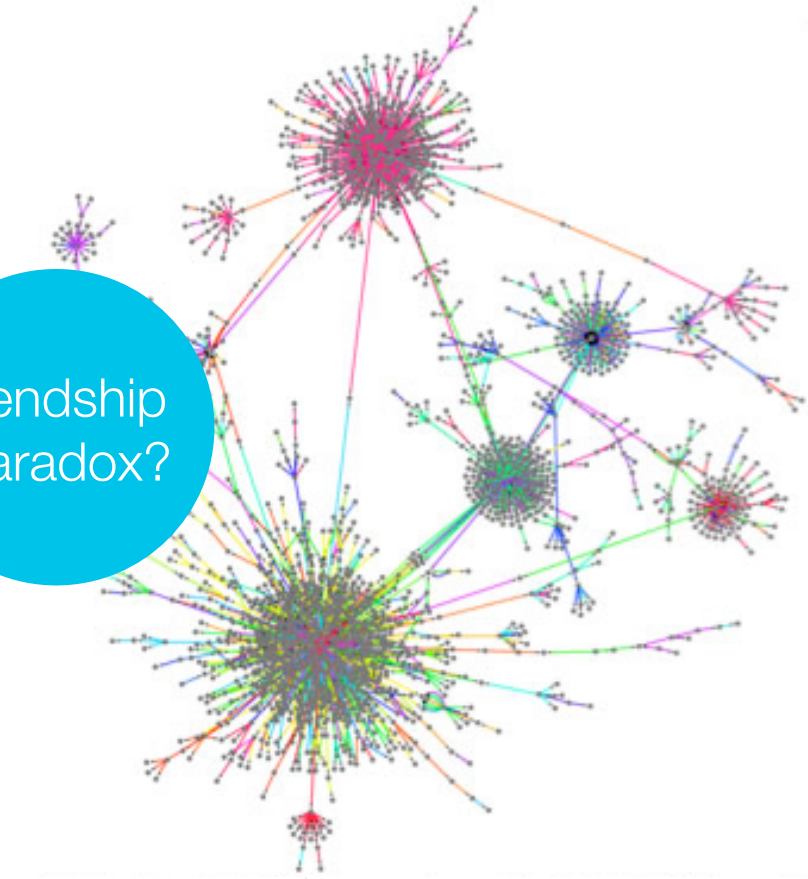
#beatcancer #twitdraw

Flu ≈ #hashtag

#pengakuan #lightupnigeria

#openwebawards

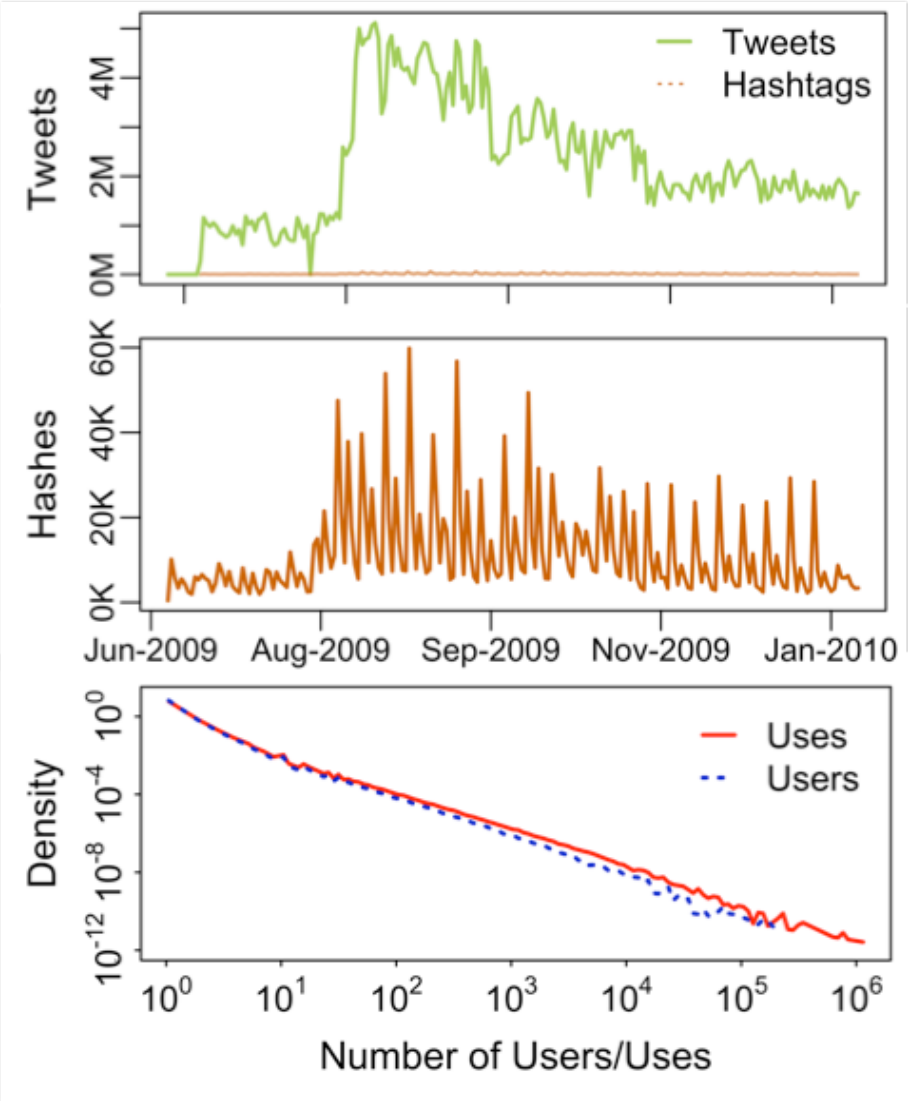
friendship paradox?



Friend ≈ Follower Followee

Detecting **global contagious outbreaks** in **Twitter**

- **Data**
- June-December 2009
- ~2/3 of Twitter
  - 476M tweets
  - 40M users
  - 1.5B follows
  - 66M tweets using a hashtag
  - 4M different hashtags
  - 1.6M users using a hashtag



3 results

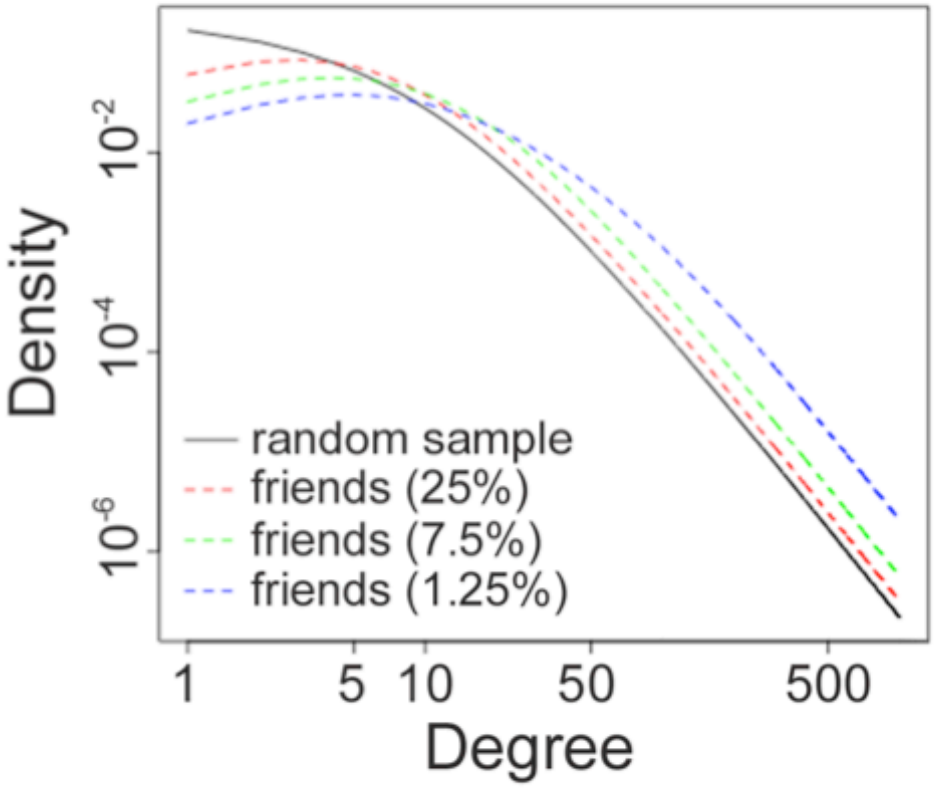
friendship paradox?

remove

mathematical tricks

and the model holds…

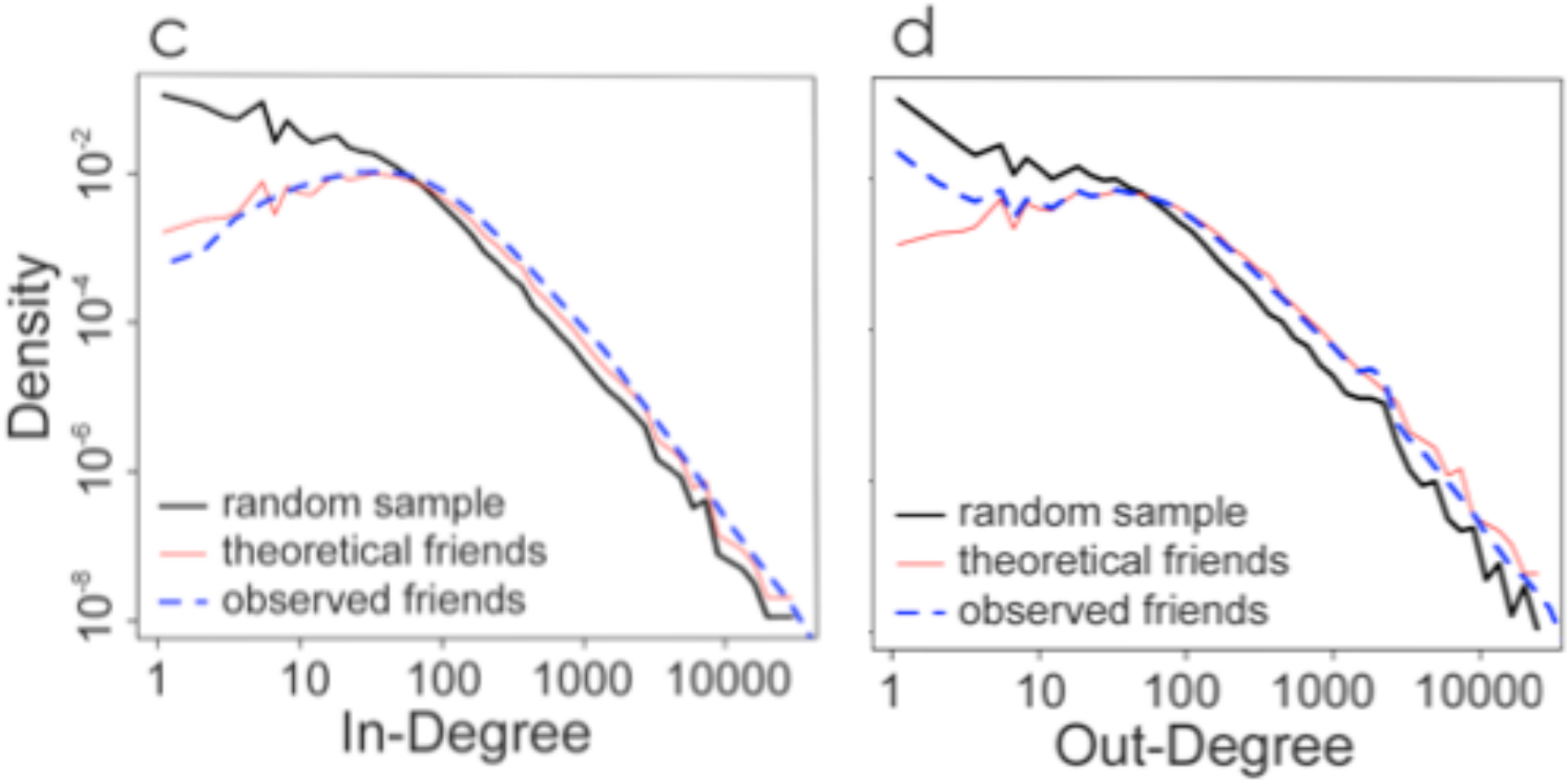
if we sample



$$\tilde{Q}(k) = \frac{\left[1-(1-\gamma)^k\right] P(k)}{A_\gamma}$$

friendship paradox?

it fits the model

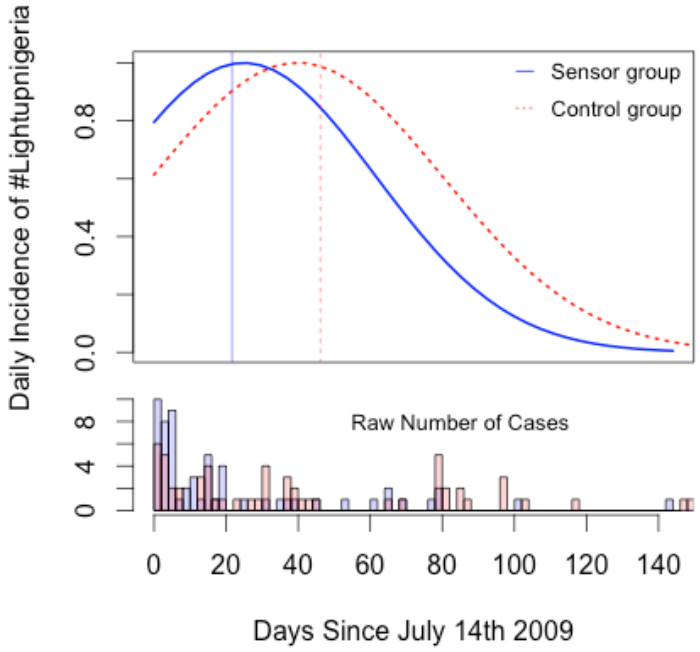


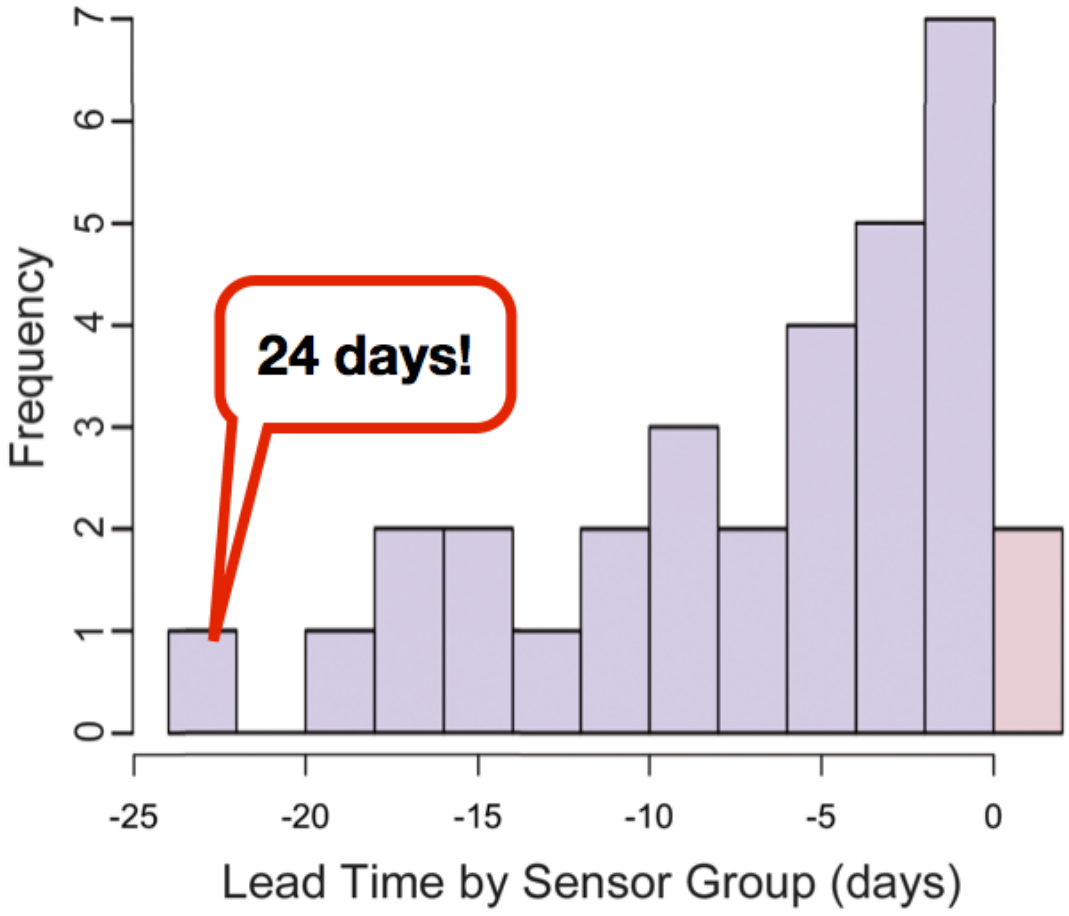
Results:

Global view (ex post)

contagious outbreak?

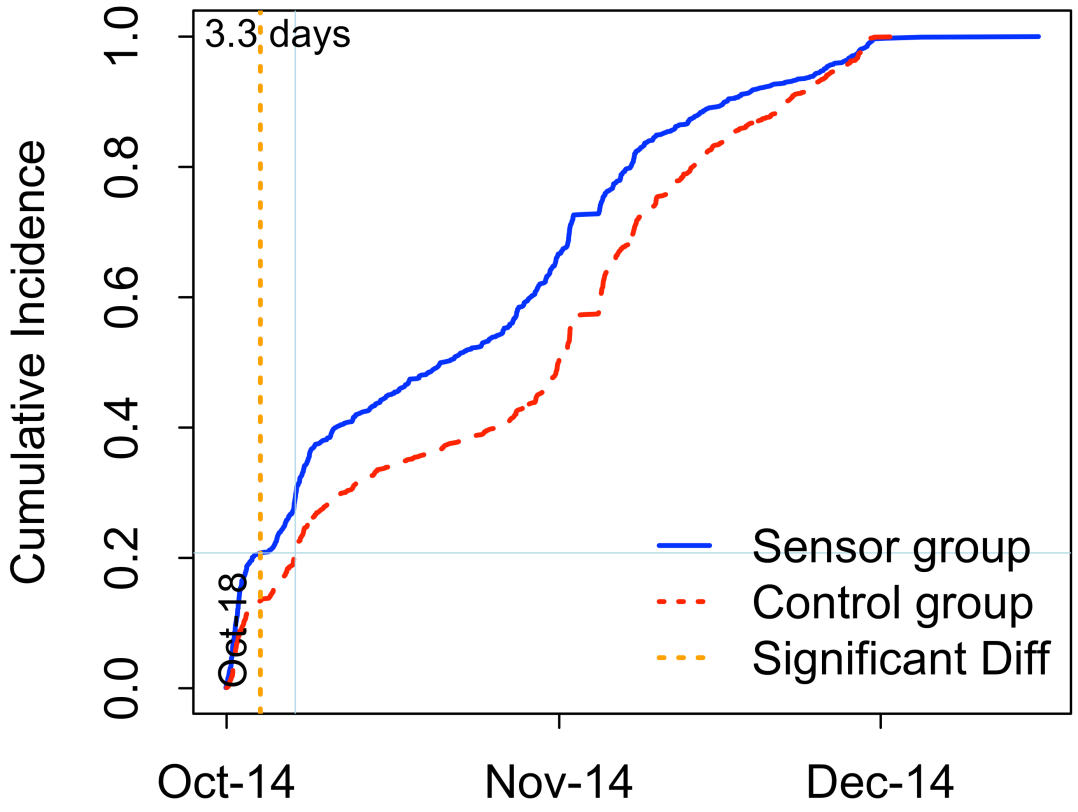
$$\langle t_{inf,i} \rangle_{i \in \text{Sensors}} \leq \langle t_{inf,i} \rangle_{i \in \text{Control}}$$

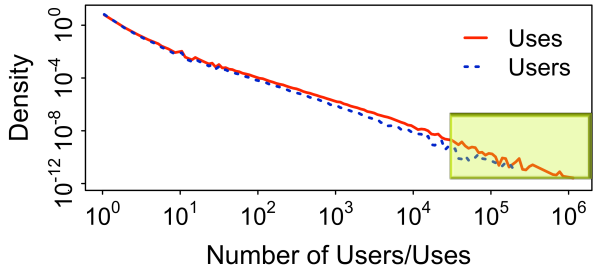


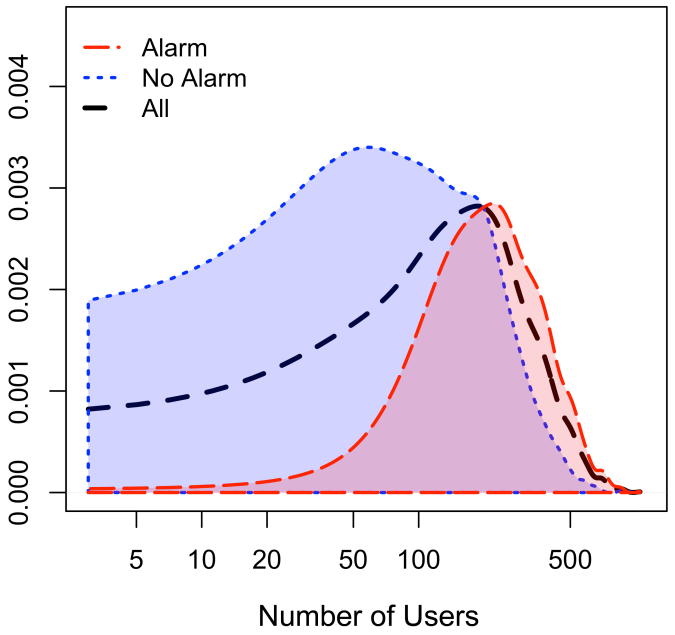


Results:

Early alarms (real time)

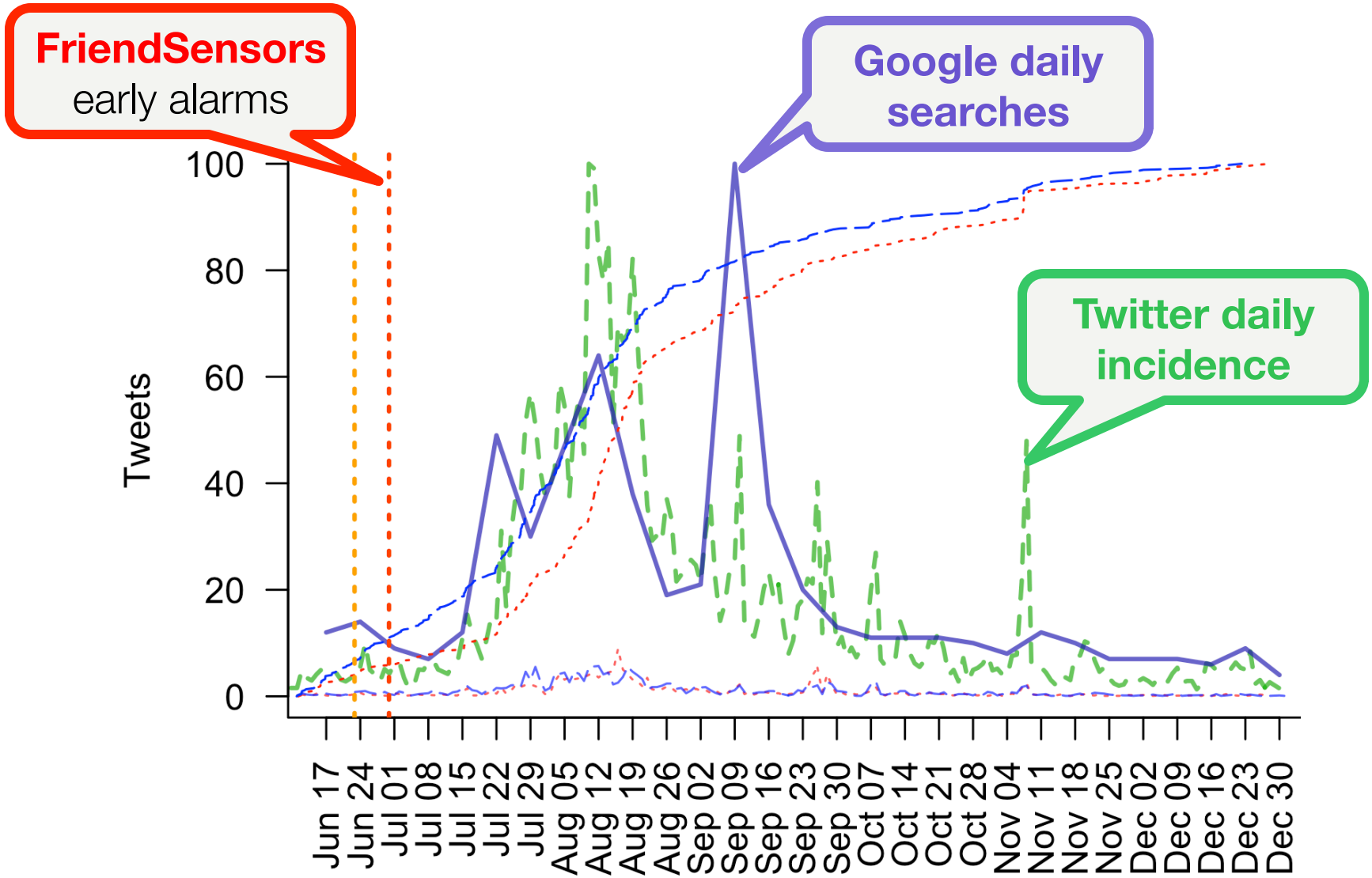






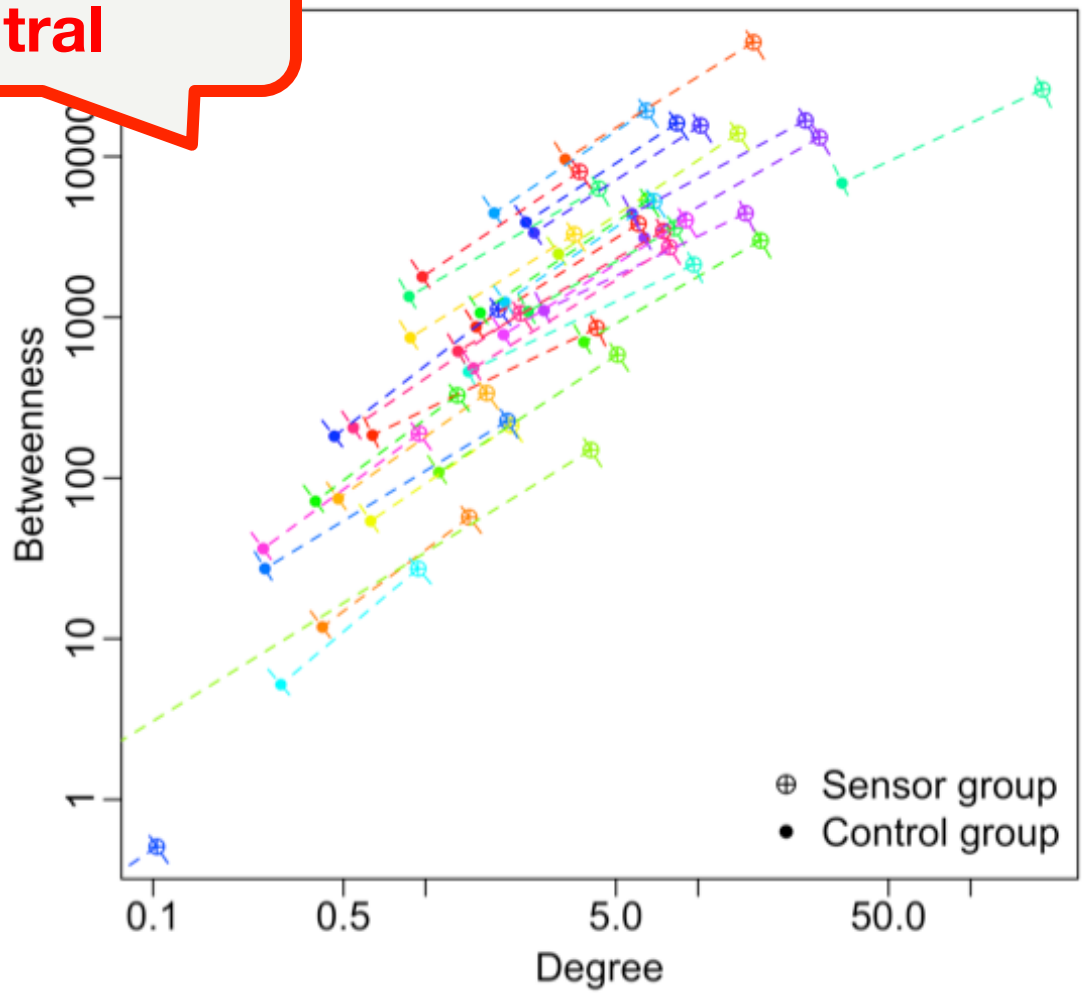
Divergence between cumulative incidence curves as early **Alarms**

Results:
how well **FriendSensors** work?



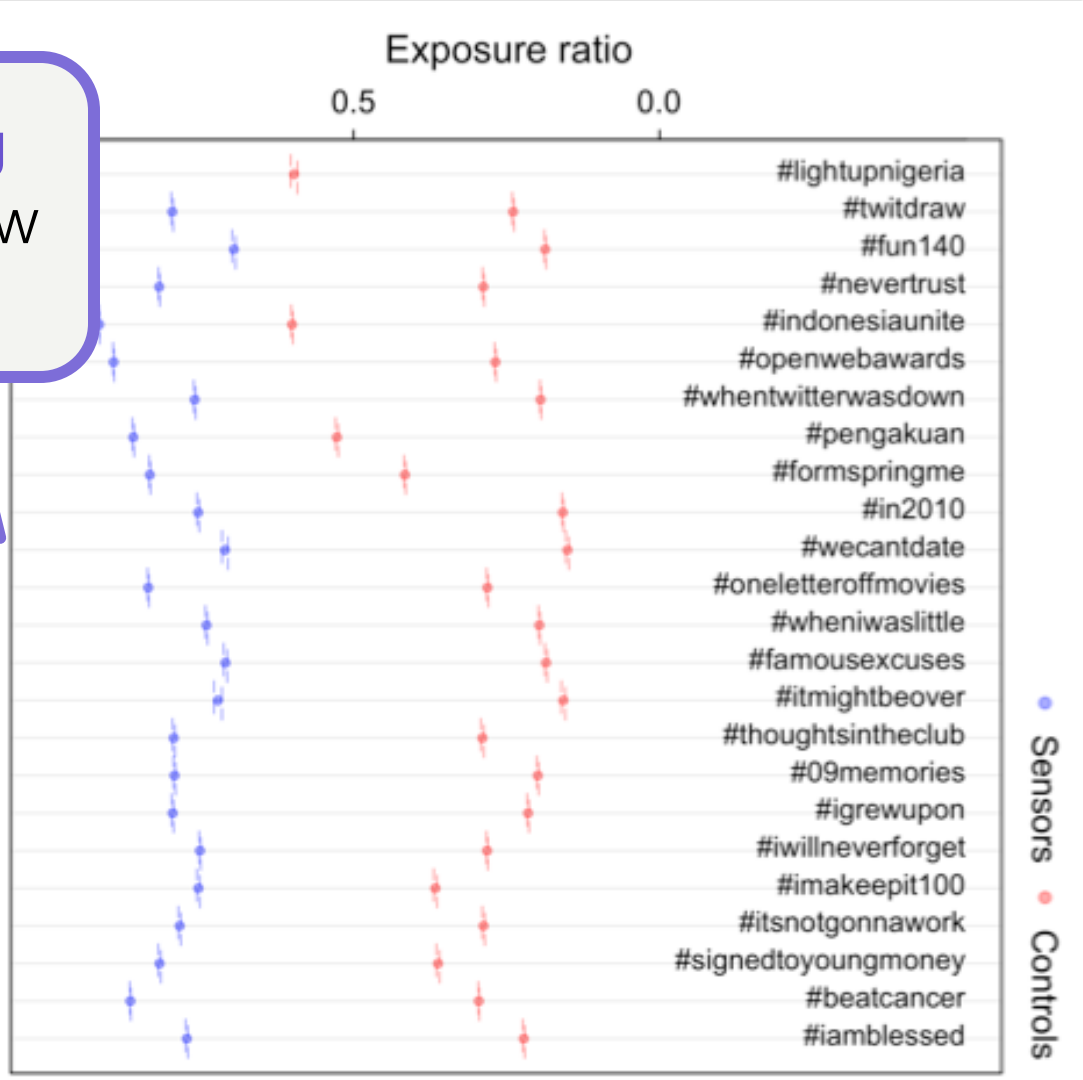
Who are the **sensors** in Twitter?

Yes, they are more **Central**



Who are the **sensors** in Twitter?

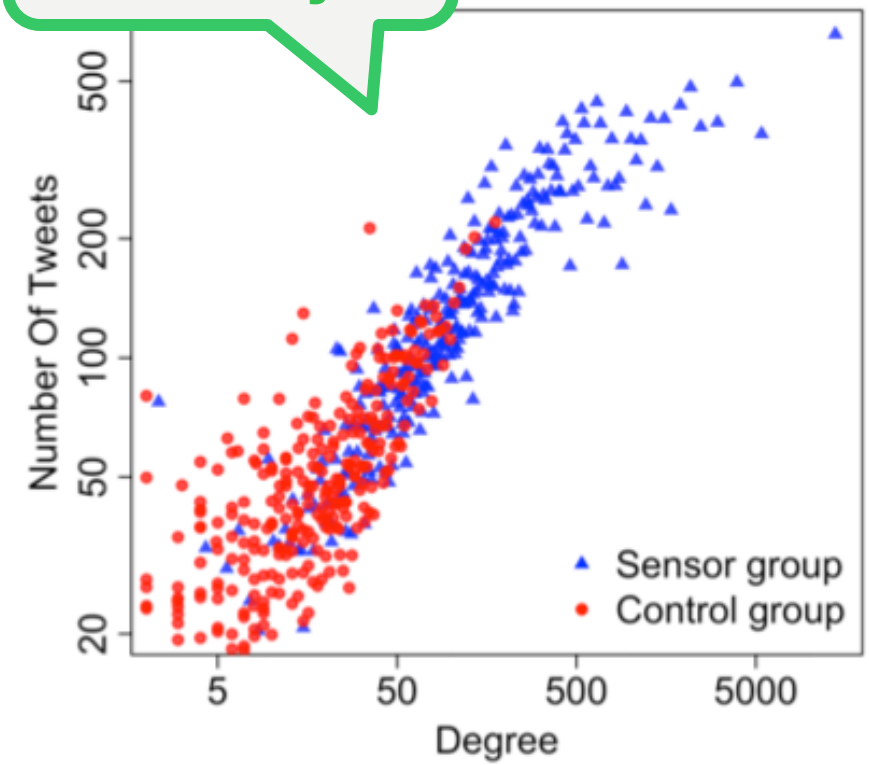
Better at **transmitting** than at **introducing** new one



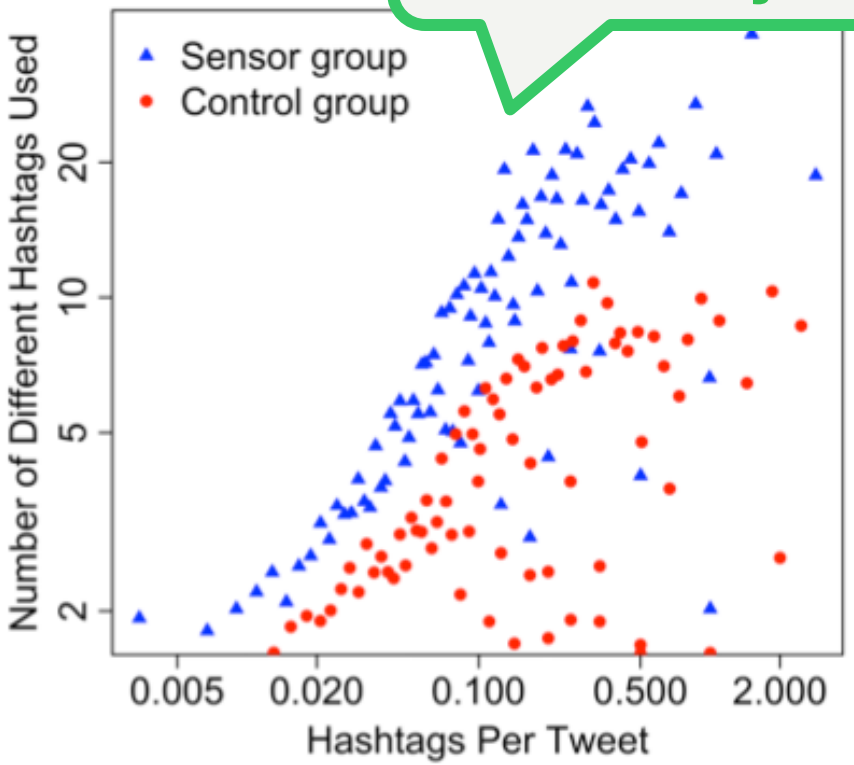
Who are the **sensors** in Twitter?

Sensors tweet more, use more hashtags, and tend to use a greater variety of hashtag

Higher **activity**



Greater **diversity**



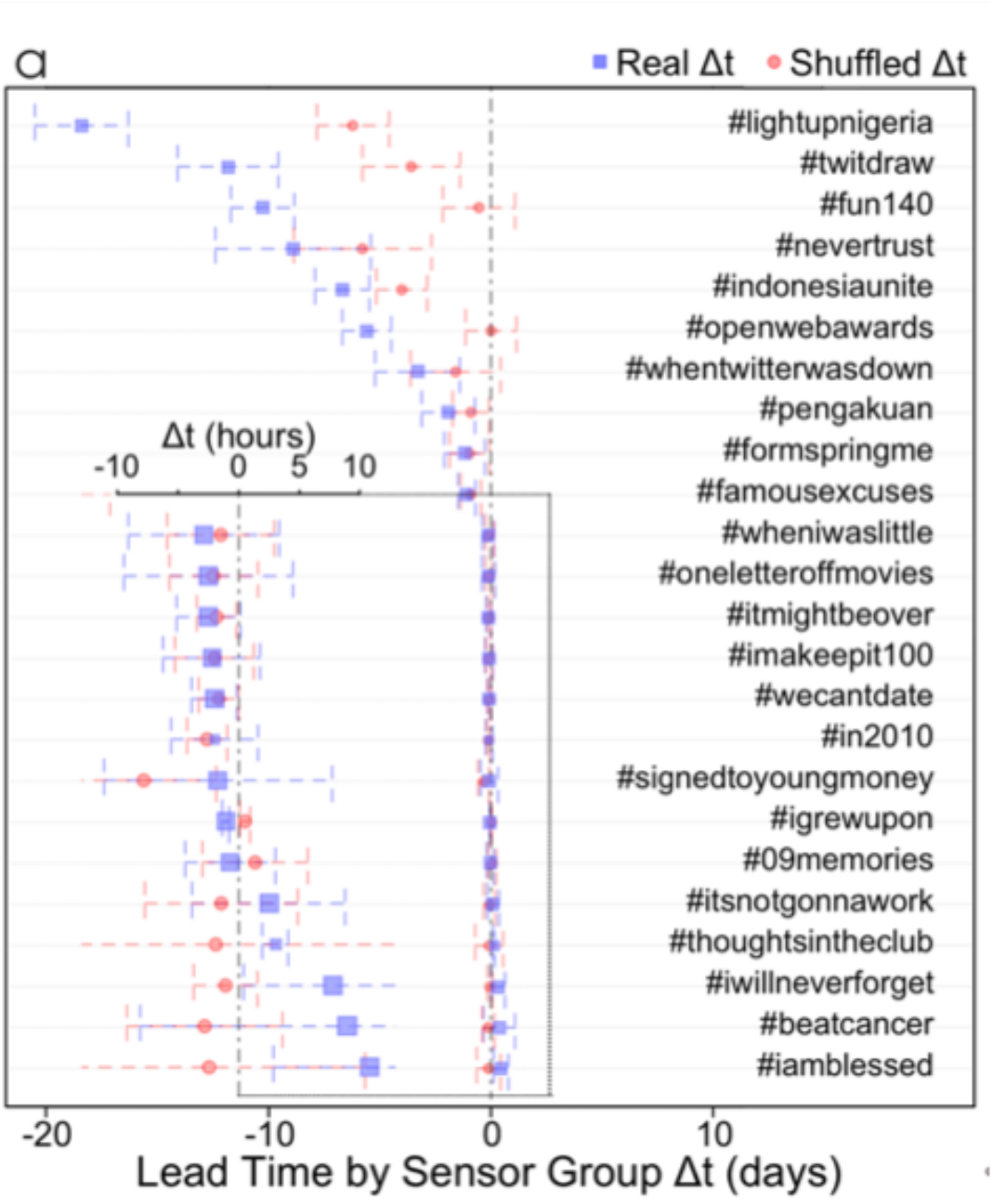
Viral or Broadcast?

Comparing the sensor and control group may help distinguishing between viral or broadcasting spreading

Real Δt is compared to the Shuffled Δt (were times of each hashtag use are shuffled)

Broadcast
Δt Real ≈ Δt Shuffled

Viral
Δt Real < Δt Shuffled



increased

centrality

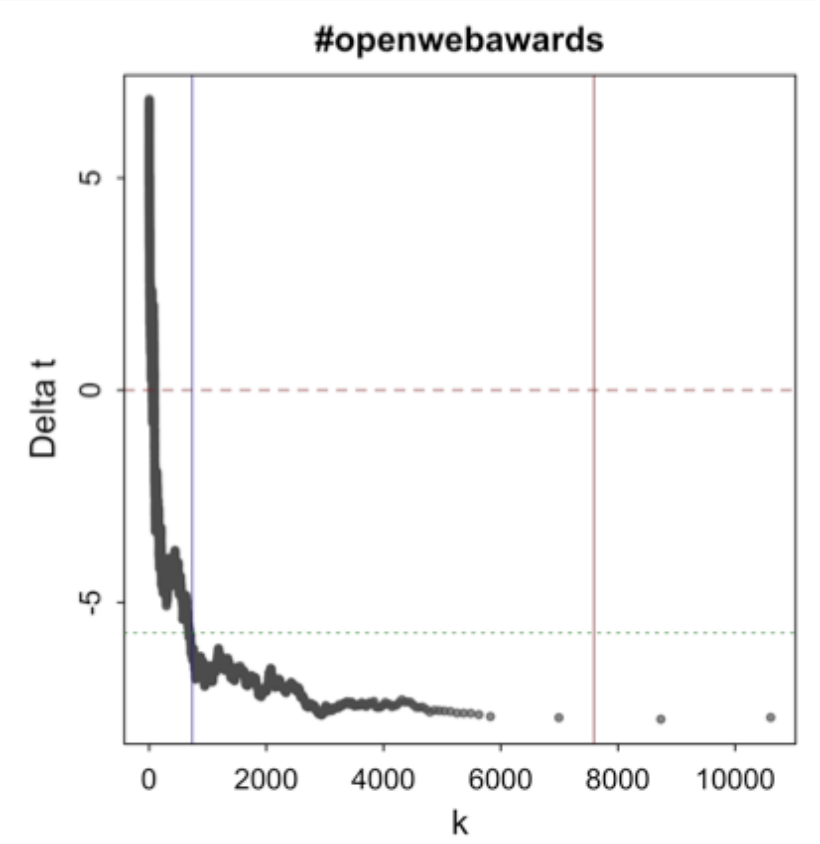
activity

diversity

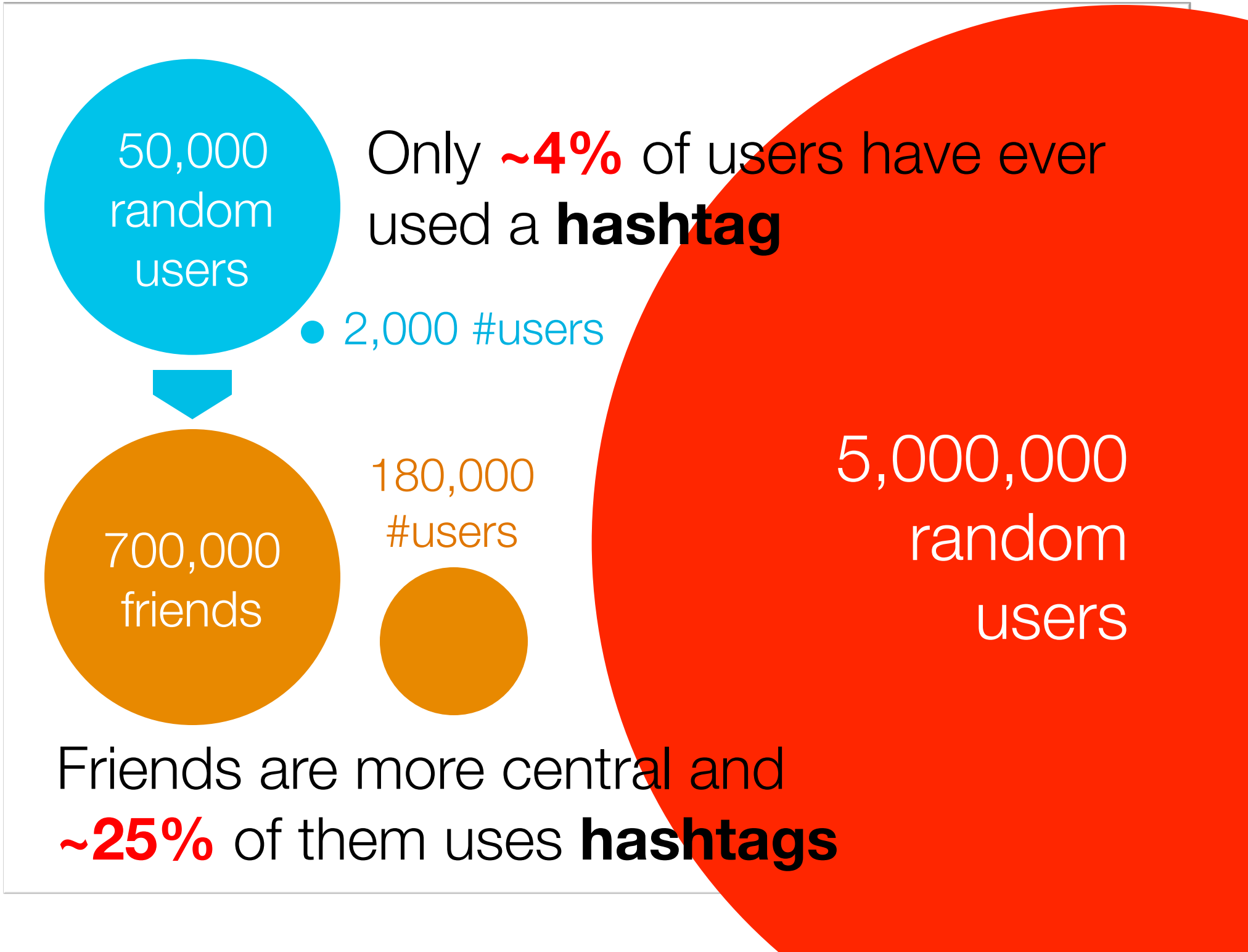
transmission

Why not **just** choose **hubs** (highly connected users)?

Yes, users with higher degree have lower (or about equal) infection time



But…


50,000
random
users
Only ~4% of users have ever
used a hashtag
2,000 #users
700,000
friends
180,000
#users
5,000,000
random
users
Friends are more central and
~25% of them uses hashtags

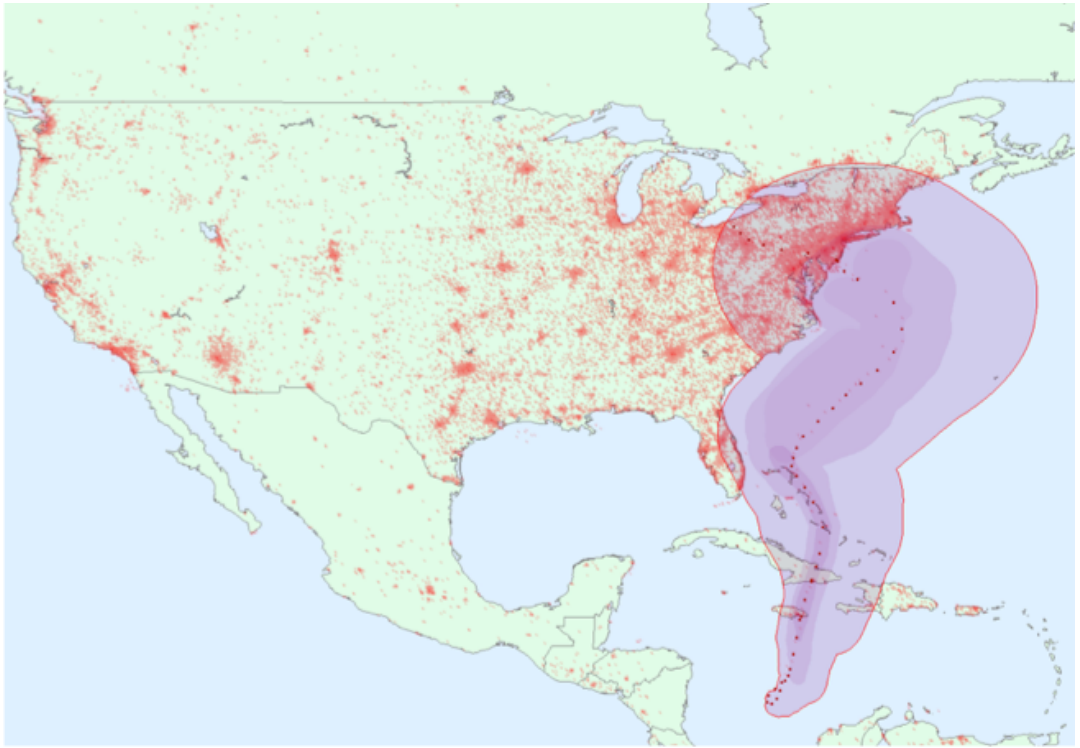
Friends vs Random

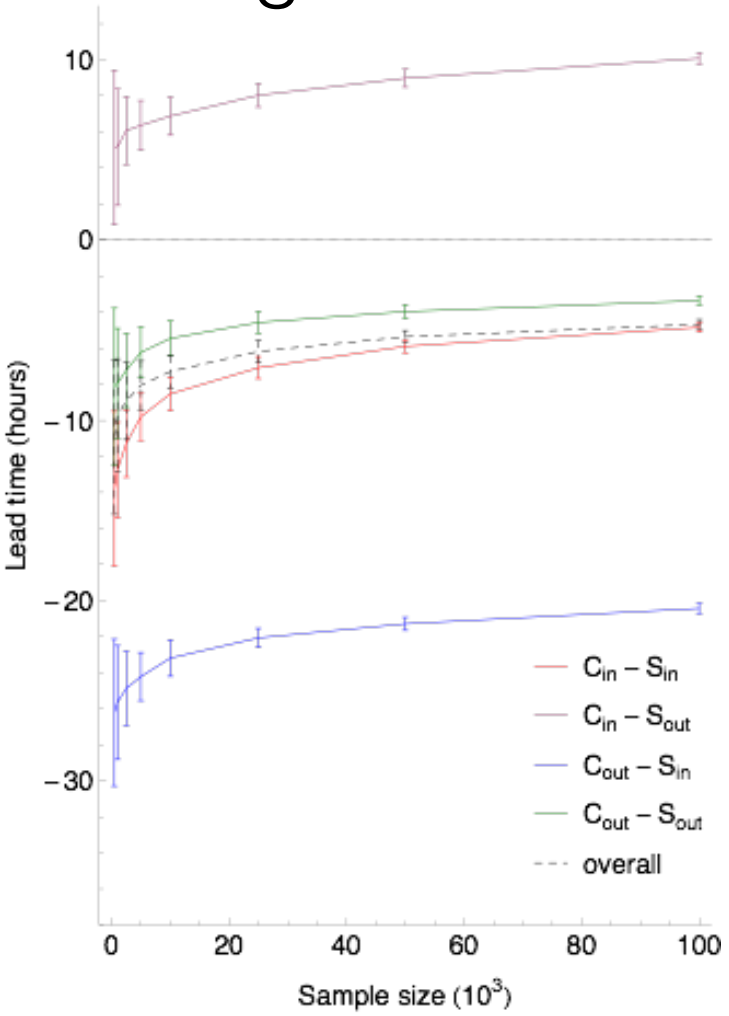
5 times **less** data to obtain a sample

6 times **more** central

4 application
to emergencies

Performance of friendsensors during **hurricane Sandy**



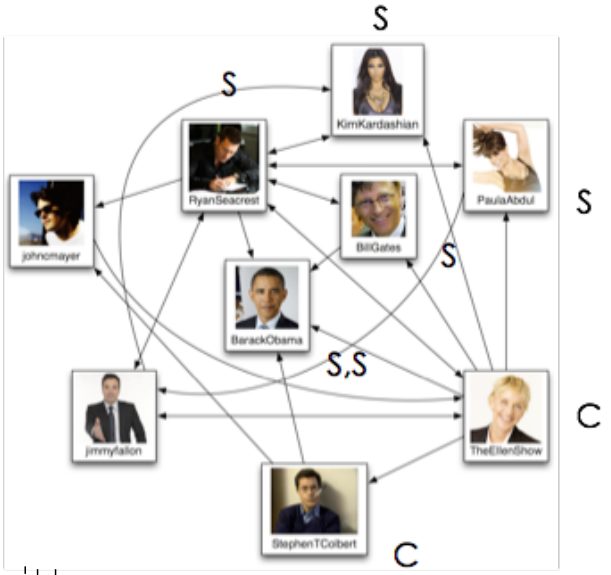


Y. Kryvasheyeou, H. Chen, E. Moro. P. van Henteryck and M. Cebrián 2014

5 conlusions

Monitoring social **BigData** requires a different approach

- Local analysis for global conclusions
- Take advantage of network structure



Our **friendsensors** method **works** on Twitter

Sensors act not only as **social hubs** (by having more connections) but also as **faster responders** (by tweeting more) and as **information hubs** (by being involved in more topics).

Difference between control and sensors can be used for **early detection**

bit.ly/friendsensors

Friendsensors: a simple yet powerful method to detect information outbreaks:

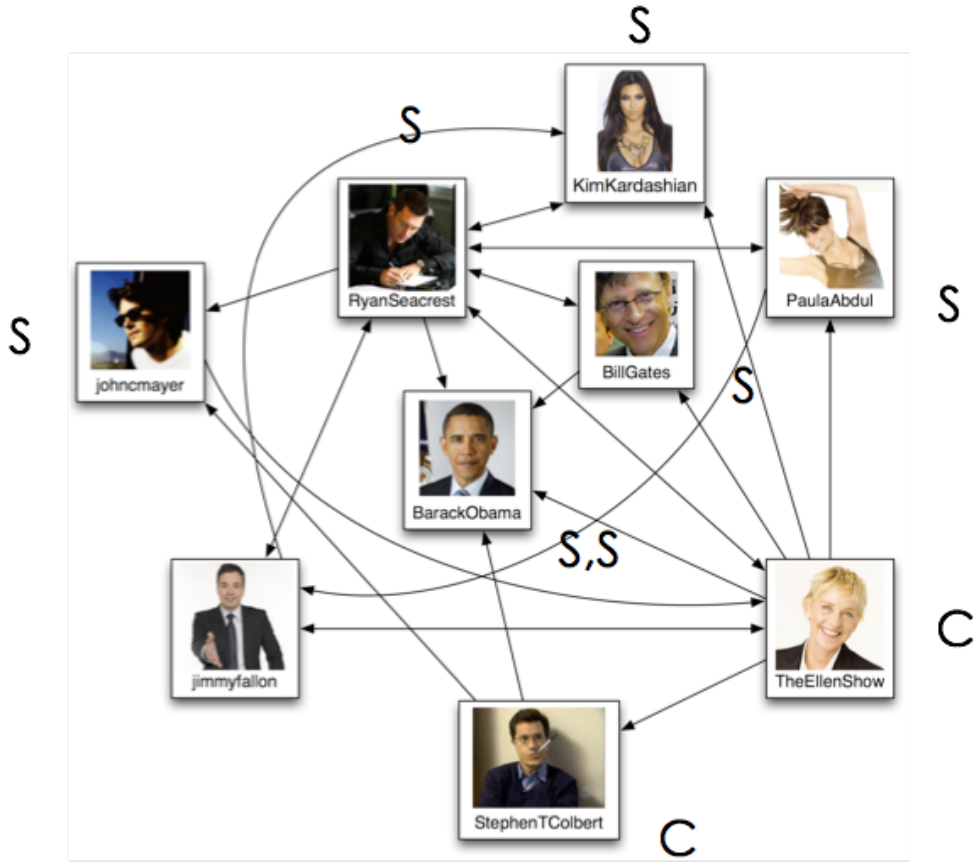
Plenty of room for improvement!
We just used the simplest way to choose sensors.

Can be used in other networks:
is based on network properties.

Can be used in highly dynamical scenarios:
choose sensors dynamicaly.

Can be adapted for geographical filters, languages, interests, etc:
choose sensors accordingly.

bit.ly/friendsensors



bit.ly/friendsensors

Manuel García-Herranz
manuelgarciaherranz@uam.es
@ranzher