Friends of friends

Sociology 204: Social Networks

Matthew J. Salganik

April 12, 2017

Logistics:

- No logisitics

Analytic Activism:
Digital Listening and the New Political Strategy

"Some of the most remarkable impacts of digital media on political activism lie not in the new types of speech it supports, but in the new forms of listening that it fosters among organized pressure groups. Organized advocacy groups are increasingly turning to digital analytics in order to gauge supporter interest, monitor public sentiment, experiment with new tactics, and craft strategies that resonate in the new media environment. In his new book, Analytic Activism, Dave Karpf discusses the heretofore overlooked role of analytics in organized political engagement. He explores how this new mode of activism works, how it is produced, what it is useful for, and what its limitations are."

David Karpf

Tuesday, April 18, 2017, 12:30 p.m., 306 Sherrerd Hall

`https://citp.princeton.edu/event/karpf/`

Favorite:

1. Strogatz (2012). Friends you can count on. New York Times.
2. Christakis (2010). How social networks predict epidemics. TED Talk.
3. Christakis and Fowler (2010). Social network sensors for early detection of contagious outbreaks. PLoS ONE.

Review:

- We can measure core discussion networks (strong ties) using name generators and name interpreters

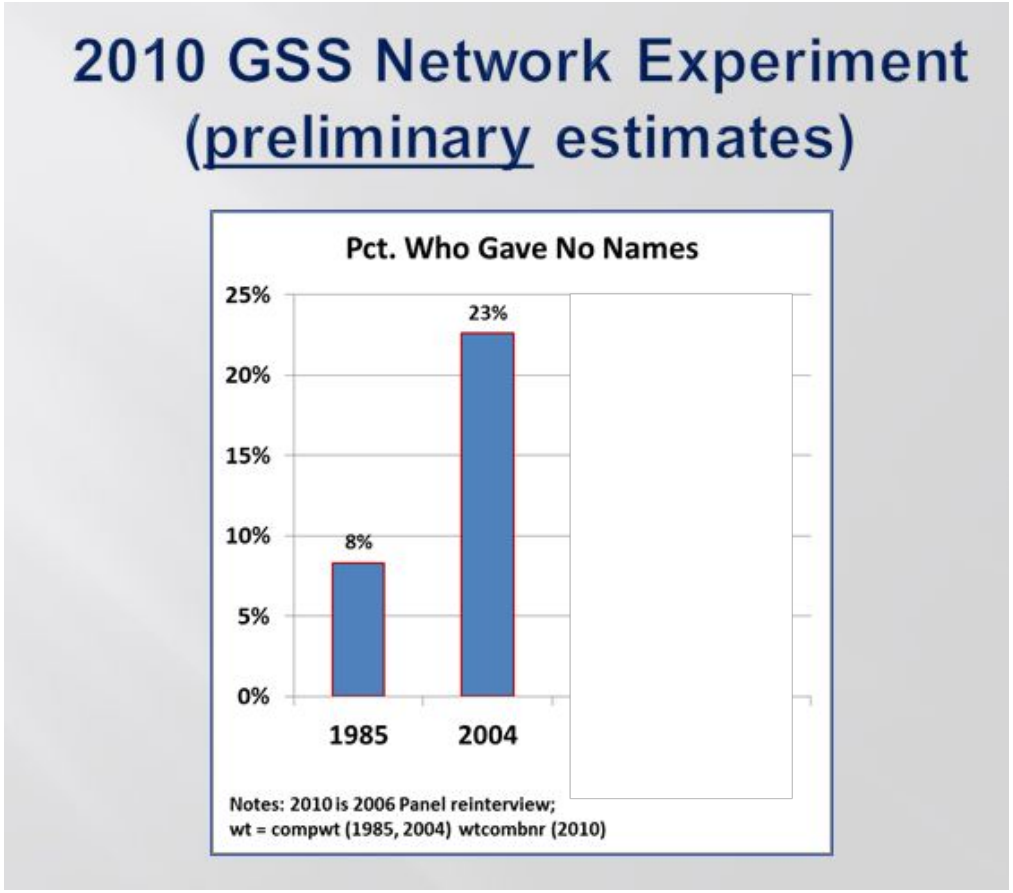
Review:

- We can measure core discussion networks (strong ties) using name generators and name interpreters
- Americans social networks appeared to be shrinking between 1985 and 2004

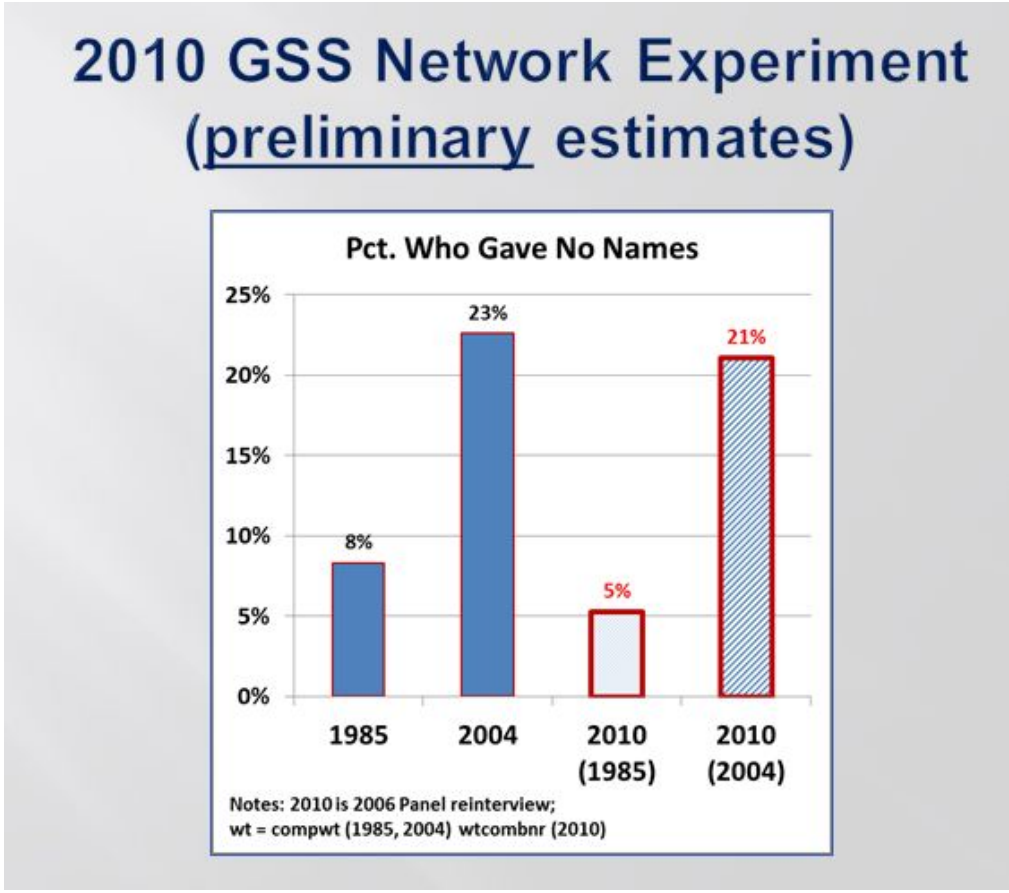
Review:

- We can measure core discussion networks (strong ties) using name generators and name interpreters
- Americans social networks appeared to be shrinking between 1985 and 2004
- That might be an artifact

After Fischer's paper there was a survey experiment:



After Fischer's paper there was a survey experiment:



Subsequent research: "Just say no" by interviewers, not respondents!

`http://dx.doi.org/10.1177%2F0003122413482919` (US)

`https://doi.org/10.1515/jos-2016-0020` (Germany)

"One lesson not to draw from this experiment is that surveys cannot be trusted. Surveys are often the only way we can learn about what people think, feel, and do. The lesson is that all studies have to be closely cross-checked and confirmed. The GSS has systematically cross-examined its methods and results over the years to establish solid findings. This experiment is another demonstration of how confidence can be built by identifying where there are errors and why. It?s way better than the impressionistic generalizing from one?s and one?s friends? experiences that underlies much of the loneliness scare."
- Claude Fischer

`https://madeinamericathebook.wordpress.com/2012/04/24/the-loneliness-scare-is-back/`

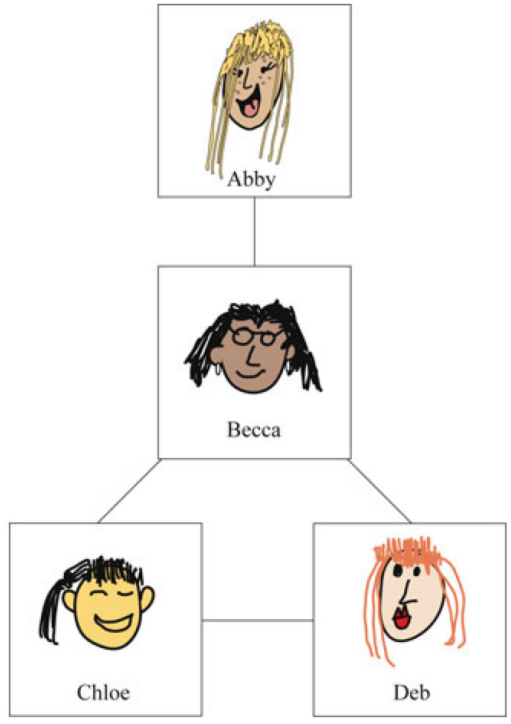
Can we learn from a study that is wrong empirically?

Can we learn from a study that is wrong empirically?
Absolutely. Process over outcome.

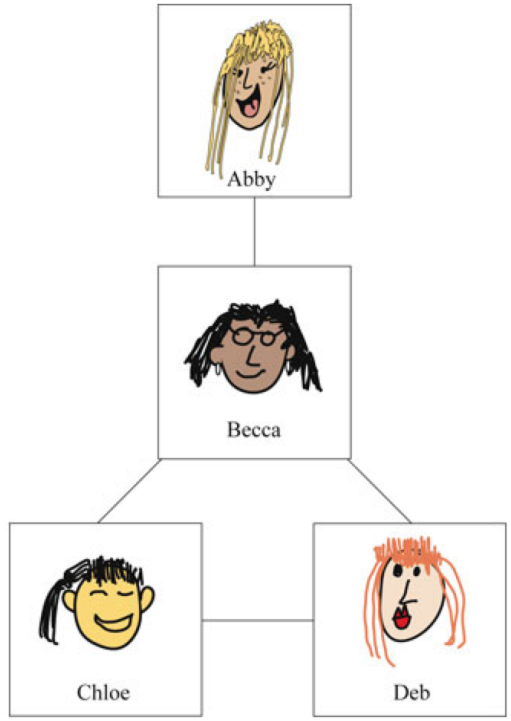
Today:

- why your friends have more friends than you
- why your friends are different in many ways
- how we can exploit that for monitoring dynamics on networks
- how we can exploit that for interventions

- Your friends probably have more friend than you
- “For any network where some people have more friends than others, it’s a theorem that the average number of friends of friends is always greater than the average number of friends of individuals.”



- Abby: Becca (3)
- Becca: Abby (1), Chloe (2), Deb (2)
- Chloe: Becca (3), Deb (2)
- Deb: Becca (3), Chloe (2)



- Abby: Becca (3)
- Becca: Abby (1), Chloe (2), Deb (2)
- Chloe: Becca (3), Deb (2)
- Deb: Becca (3), Chloe (2)

- Avg number of friends of individuals: 2
- Avg number of friends of friends: 2.25

In which situation will there be more of difference between the average number of friends of friends and the average number of friends of individuals?

1. power-law degree distribution
2. Poisson degree distribution

In which situation will there be more of difference between the average number of friends of friends and the average number of friends of individuals?

1. power-law degree distribution
2. Poisson degree distribution

mean number of friends of friends = mean(d) + variance(d)/mean(d)

Imagine that next year humans discover that there is life on Mars. It was discovered that Martians have a special relationship called occra, which is similar to friendship in human, but the precise characteristics of this relationship are still be studied. However, one thing that has been discovered is that occra is an undirected relationship; that is, you would say that Peter and Paul are occras not that Peter is Paul's occra. If not all Martians have the same number of occras, which of the following must be true about the network of occras between Martians? Please do not assume that Martian social relations look anything like human social relations.

1. There is short average path length
2. There is high clustering coefficient
3. both (1) and (2)
4. The mean number of occras of occras will be greater than or equal to the mean number of occras of Martians
5. We don't know for sure any properties of the occras network because we don't know enough about this type of relationship

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

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

http://www.jstor.org/stable/2781907

Implication of this pattern: Social distortion

Implication of this pattern: Social distortion
Any trait that is correlated with degree is more likely to be over-estimated

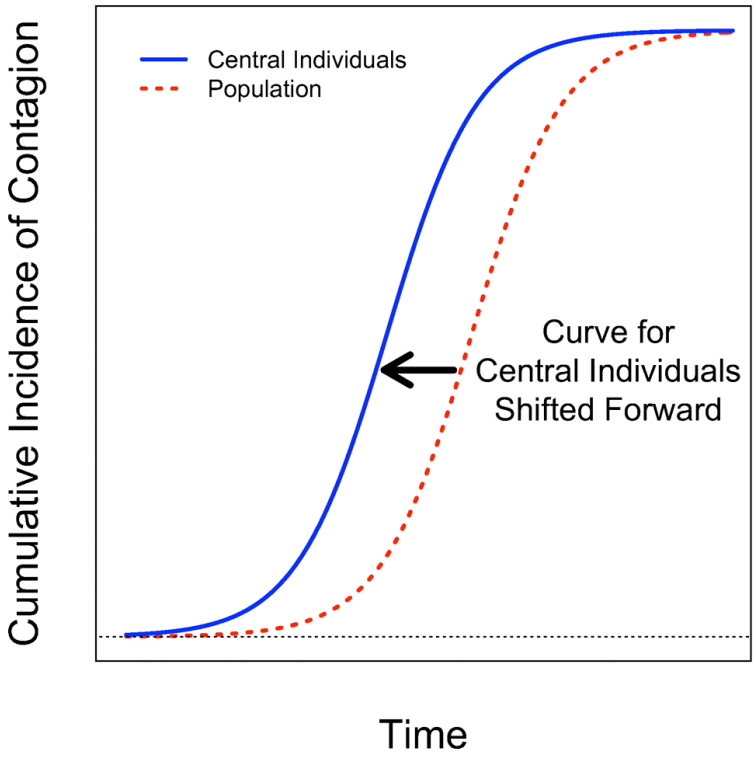
Exploiting this pattern

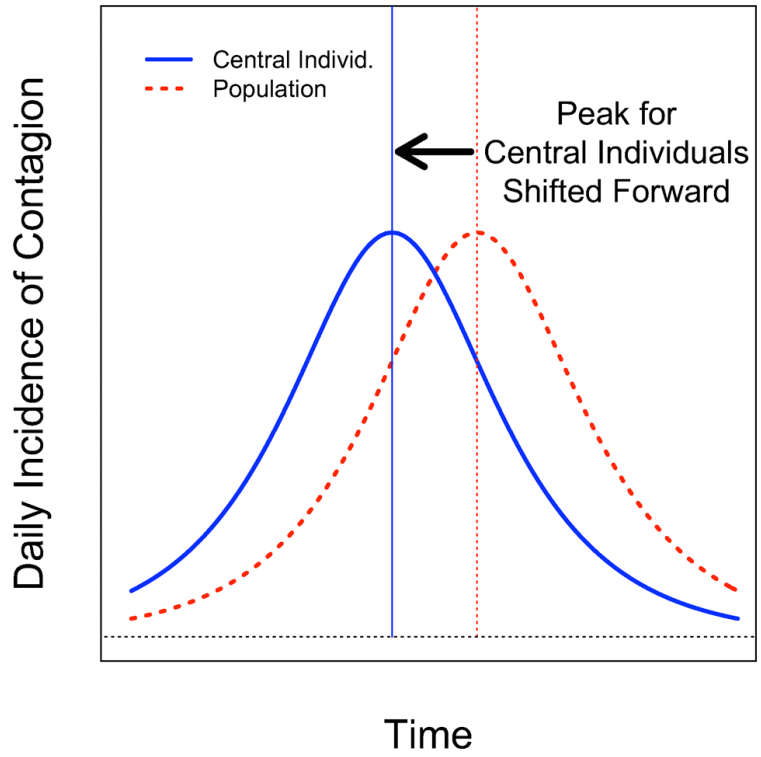
3 methods of tracking influenza:

- CDC Method (measure the past)
- Google Flu Method (measure the present)
- Social sensors (predict the future)

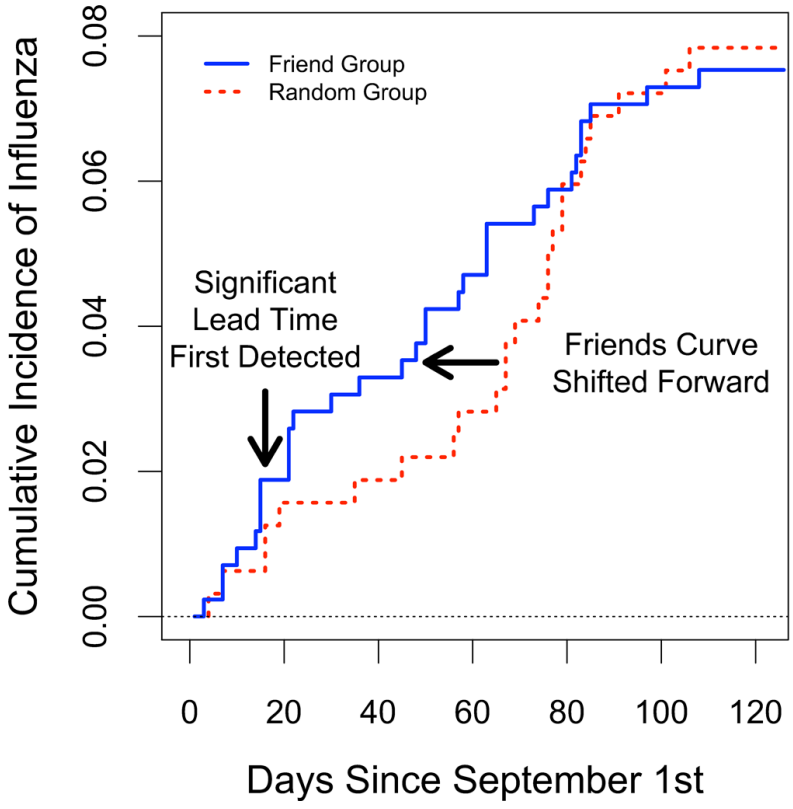
1. Randomly sample Harvard students: Random group (n=319)
2. Ask them to refer friends: Friend group (n=425)

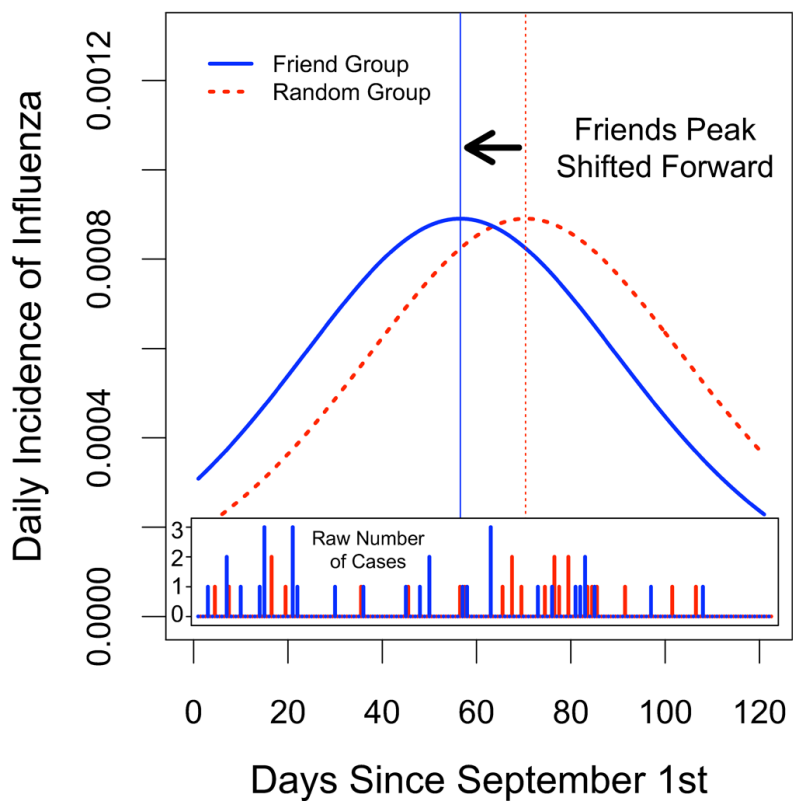
Prediction:

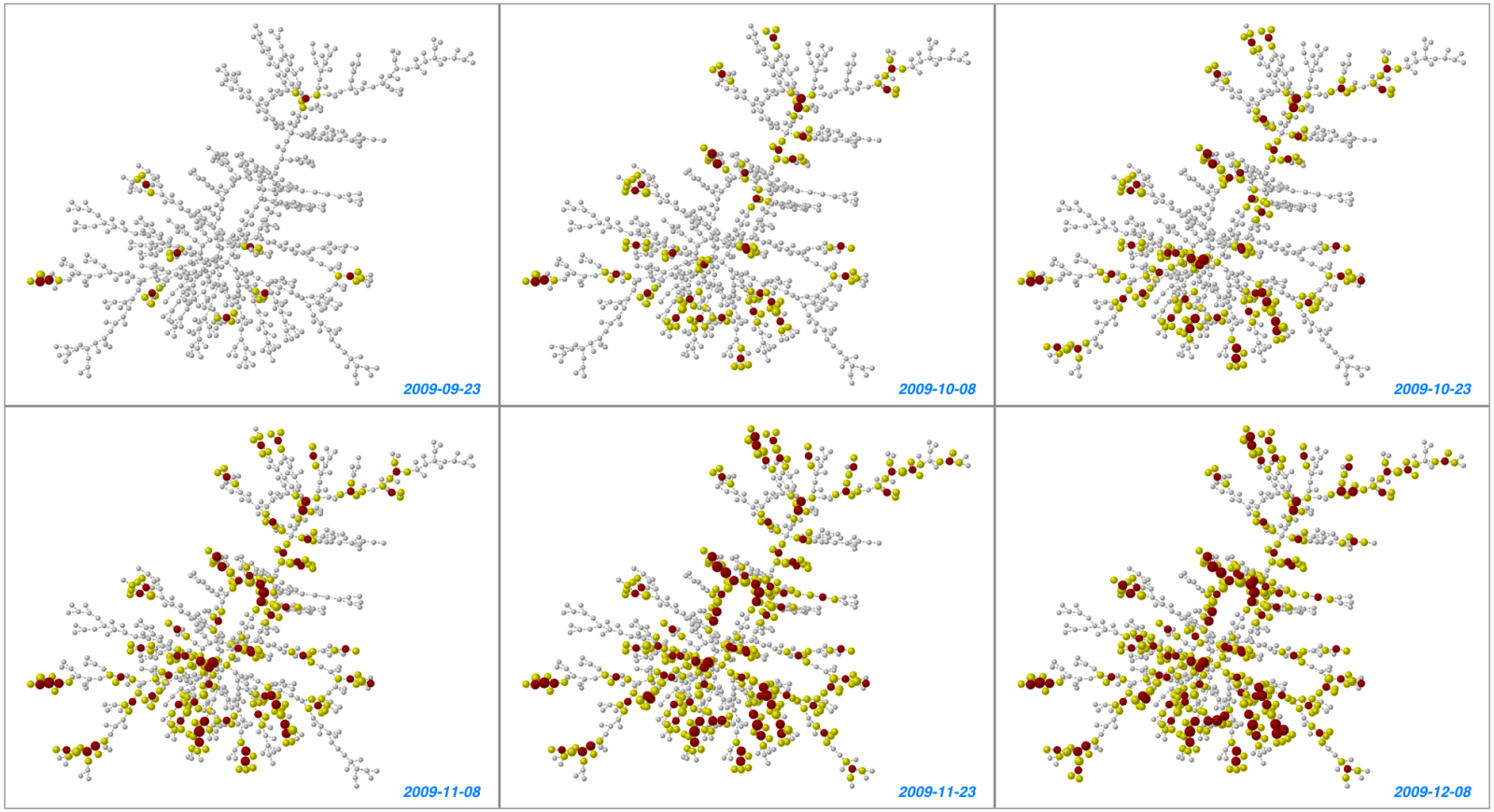

Central Individuals
Population
Curve for
Central Individuals
Shifted Forward
Cumulative Incidence of Contagion
Time

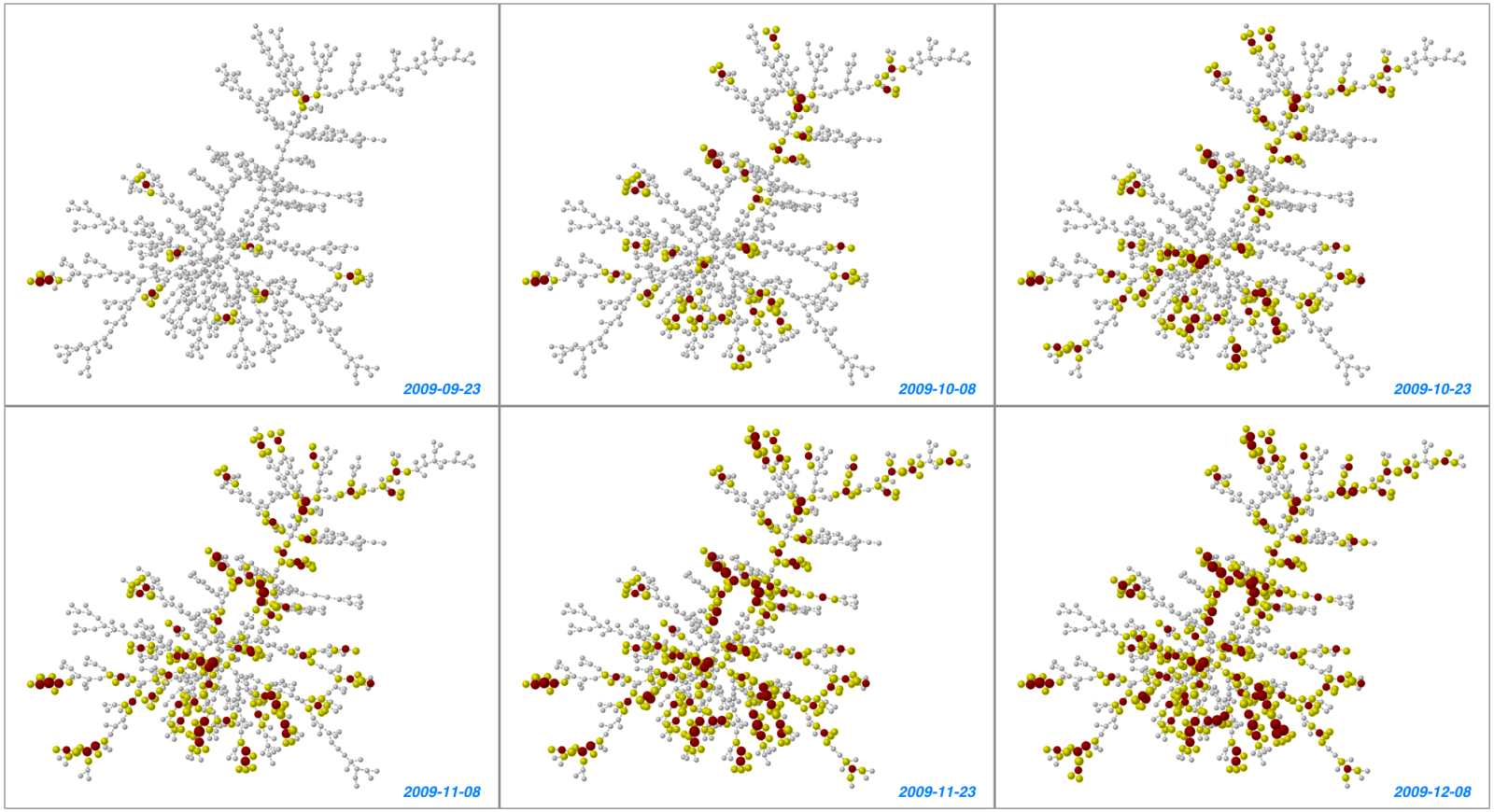

Central Individ.
Population
Peak for
Central Individuals
Shifted Forward
Daily Incidence of Contagion
Time

Prediction:


Cumulative Incidence of Influenza
0.00
0.02
0.04
0.06
0.08
Friend Group
Random Group
Significant
Lead Time
First Detected
Friends Curve
Shifted Forward
0
20
40
60
80
100
120
Days Since September 1st

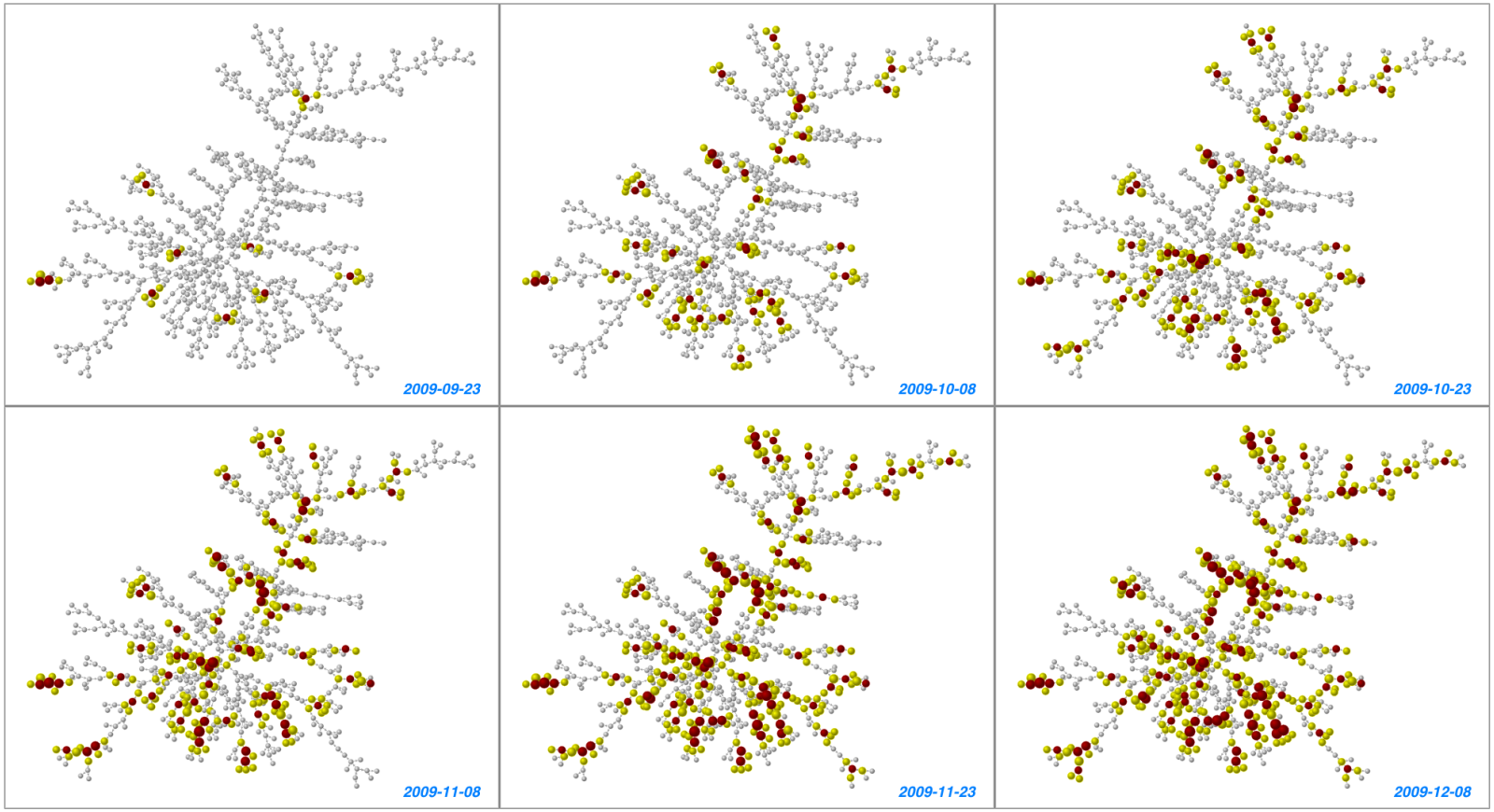

Daily Incidence of Influenza
0.0000
0.0004
0.0008
0.0012
Friend Group
Random Group
Friends Peak
Shifted Forward
Raw Number
of Cases
3
2
1
0
0
20
40
60
80
100
120
Days Since September 1st


2009-09-23
2009-10-08
2009-10-23
2009-11-08
2009-11-23
2009-12-08



Caveats:

- “centrality” vs high degree



Caveats:

- "centrality" vs high degree
- poor representation of Harvard undergrad network

"anything that spreads via a form of social contagion" For what kinds of spreading might high degree people not get infected early?

"anything that spreads via a form of social contagion" For what kinds of spreading might high degree people not get infected early?

threshold behavior; it takes more people to trigger high degree people

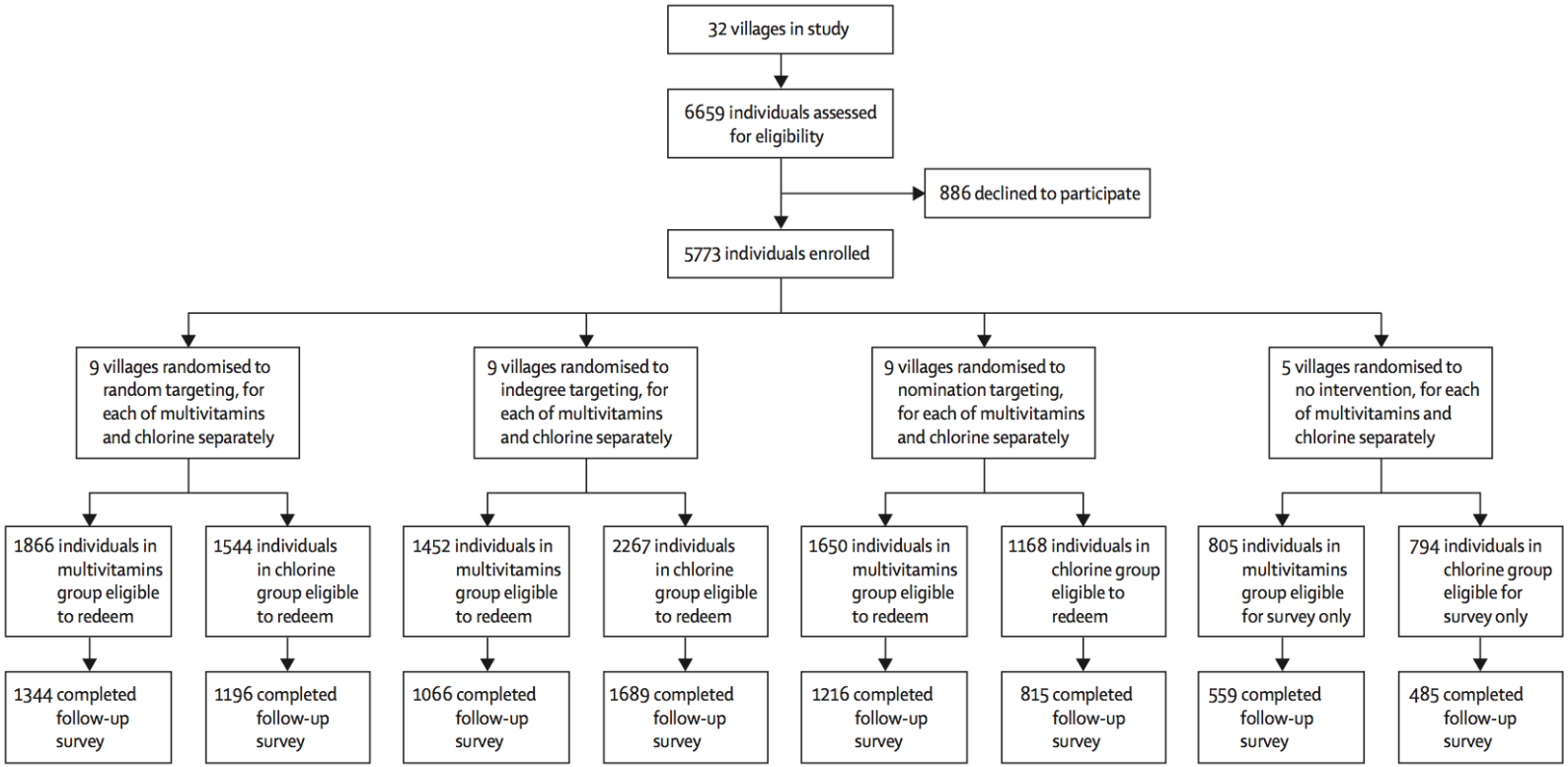
Monitoring to intervention

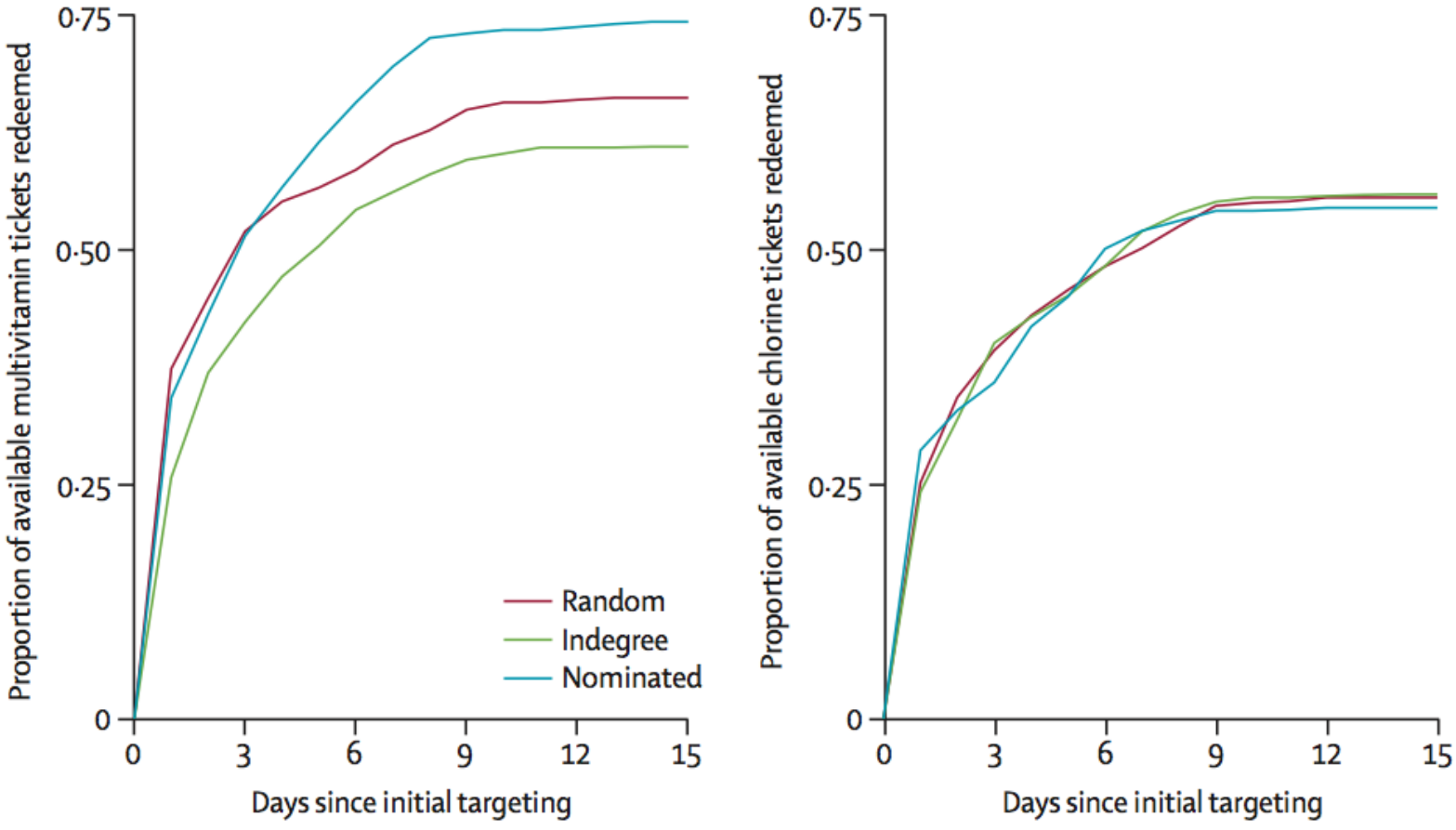
Social network targeting to maximise population behaviour change: a cluster randomised controlled trial

David A Kim, Alison R Hwong, Derek Stafford, D Alex Hughes, A James O'Malley, James H Fowler, Nicholas A Christakis

Approximate design:

- Map networks in 32 villages
- Randomize four targeting strategies: random, highest degree, friend of friend, none
- Two outcomes: multivitamin usage, chlorinated water


32 villages in study
6659 individuals assessed for eligibility
886 declined to participate
5773 individuals enrolled
9 villages randomised to random targeting, for each of multivitamins and chlorine separately
9 villages randomised to indegree targeting, for each of multivitamins and chlorine separately
9 villages randomised to nomination targeting, for each of multivitamins and chlorine separately
5 villages randomised to no intervention, for each of multivitamins and chlorine separately
1866 individuals in multivitamins group eligible to redeem
1544 individuals in chlorine group eligible to redeem
1452 individuals in multivitamins group eligible to redeem
2267 individuals in chlorine group eligible to redeem
1650 individuals in multivitamins group eligible to redeem
1168 individuals in chlorine group eligible to redeem
805 individuals in multivitamins group eligible for survey only
794 individuals in chlorine group eligible for survey only
1344 completed follow-up survey
1196 completed follow-up survey
1066 completed follow-up survey
1689 completed follow-up survey
1216 completed follow-up survey
815 completed follow-up survey
559 completed follow-up survey
485 completed follow-up survey


0·75
0·50
0·25
0
Proportion of available multivitamin tickets redeemed
0
3
6
9
12
15
Days since initial targeting
Random
Indegree
Nominated
0·75
0·50
0·25
0
Proportion of available chlorine tickets redeemed
0
3
6
9
12
15
Days since initial targeting

To learn more:

- Press release: `http://news.yale.edu/2015/05/05/social-network-experiments-create-tipping-point-improv`
- Commentary paper: `http://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)60503-7/abstract`
- Paper: `https://www.ncbi.nlm.nih.gov/pubmed/25952354`

Key points:

- your friends probably have more friends than you
- sample of friends can behave different from a sample of people

What's next?

What's next?

Networks and hidden populations at risk for HIV/AIDS

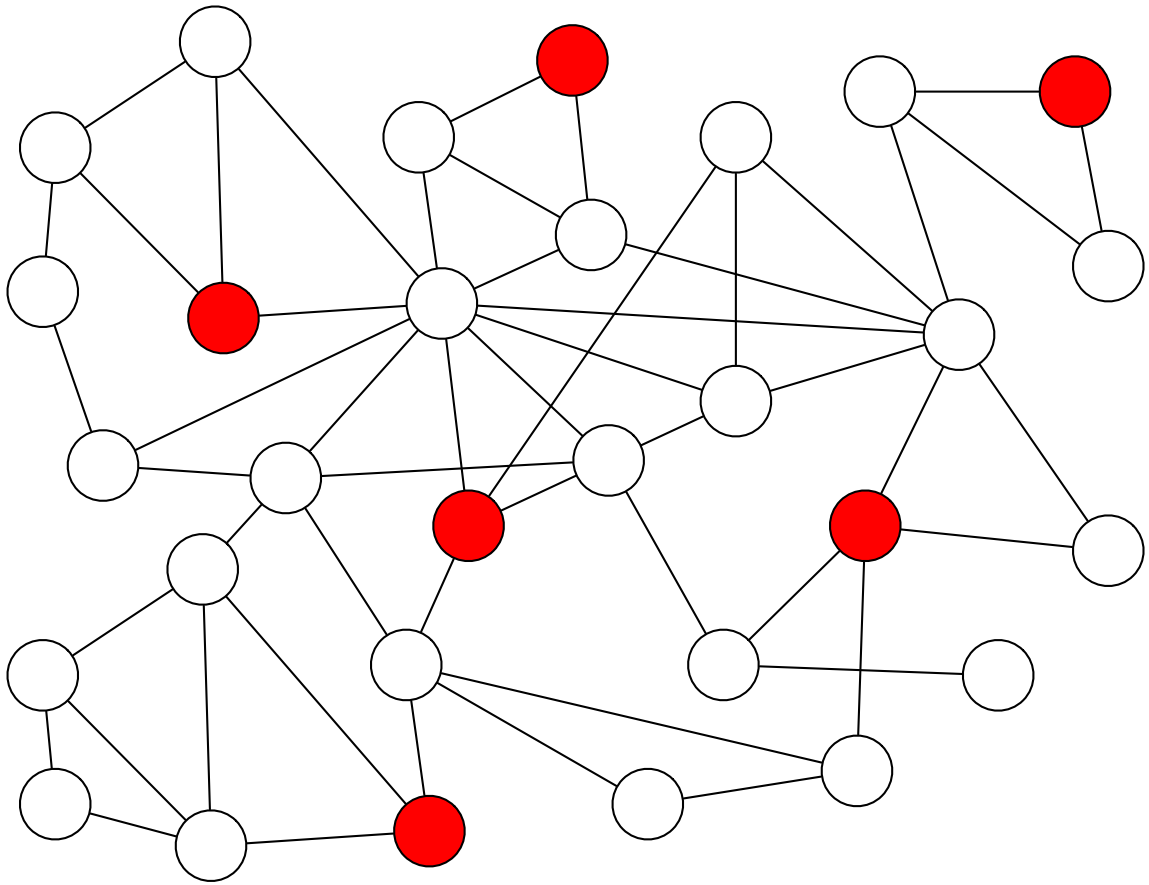
- Bernard, H.R. et al. (2010). Counting hard-to-count populations: The network scale-up method for public health. Sexually Transmitted Infections.
- Feehan et al. (2016). Quality vs. Quantity: A survey experiment to improve the network scale-up method. American Journal of Epidemiology.

There are between 31 million and 35 million people worldwide living with HIV/AIDS. In most countries, the disease is concentrated in three high-risk groups:

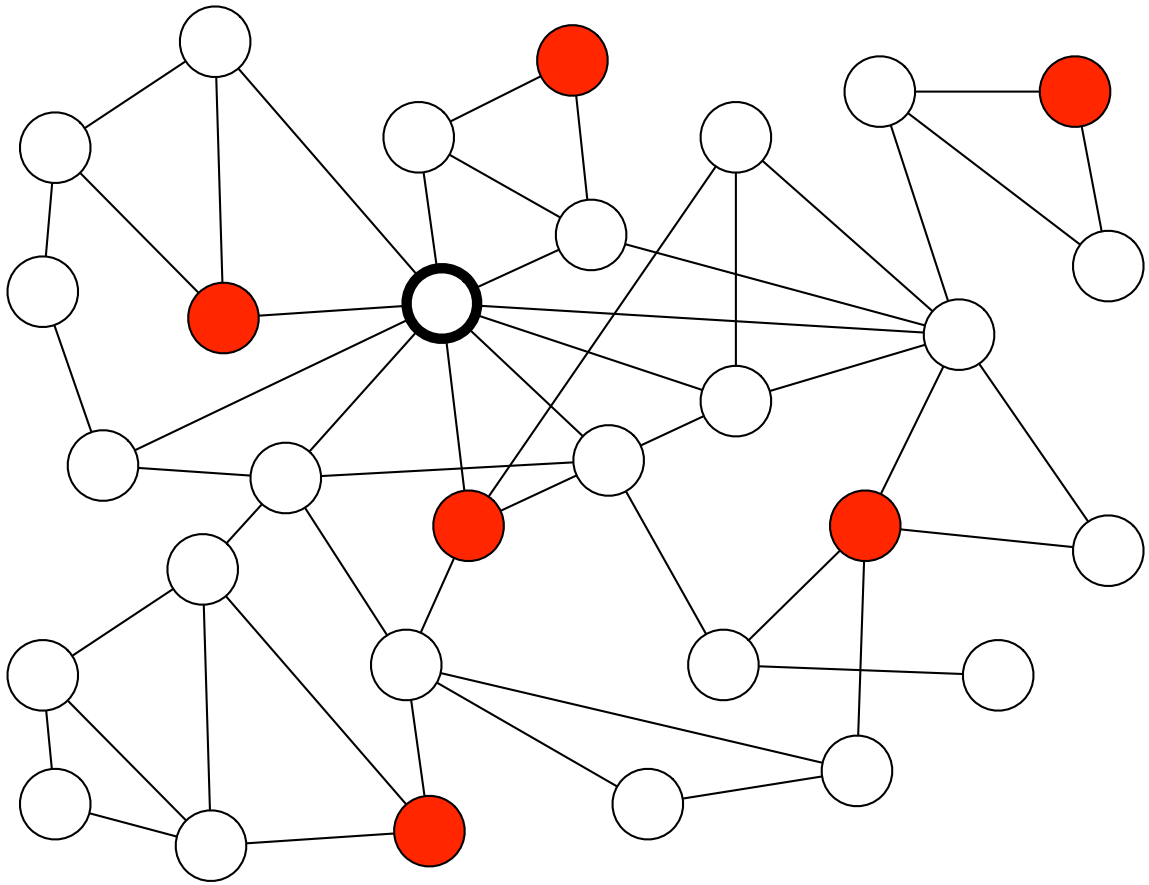
- injection drug users
- commercial sex workers
- men who have sex with men

Better information about these group can be used to understand and control the spread of HIV: “know your epidemic”

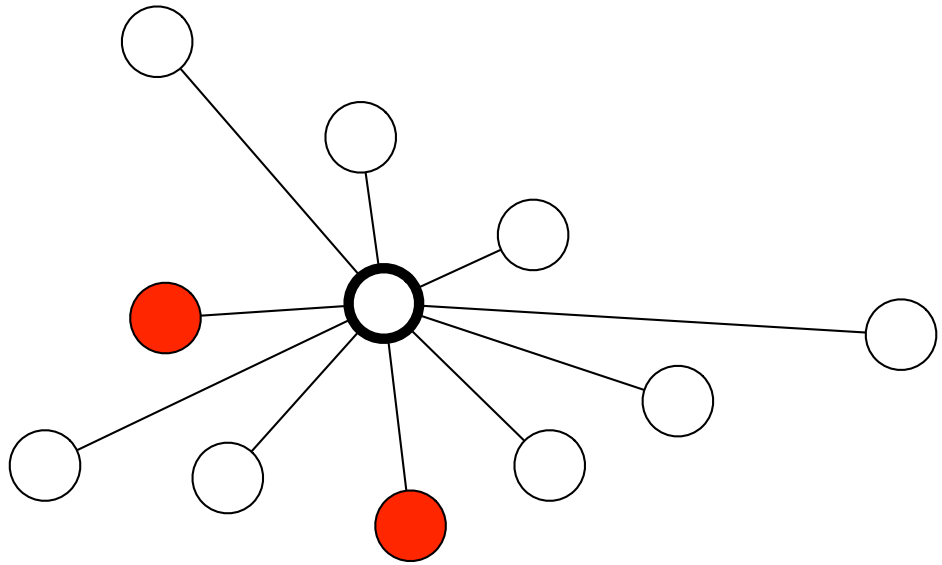
Network scale-up method



Network scale-up method



Network scale-up method



$$\hat{N}_H = \frac{2}{10} \times 30 = 6$$

If $\underbrace{y_{i,k} \sim Bin(d_i, N_k/N)}_{\text{basic scale-up model}}$, then maximum likelihood estimator is

$$\hat{N}_H = \frac{\sum_i y_{i,H}}{\sum_i \hat{d}_i} \times N$$

- $\hat{N}_H$: number of people in the hidden population
- $y_{i,H}$: number of people in hidden population known by person i
- $\hat{d}_i$: estimated number of people known by person i
- N: number of people in the population

See Killworth et al., (1998)

- Requires a random sample from the entire population
- Respondents are asked:
  - How many people do you know who are drug injectors?
  - How many women do you know that have given birth in the last 12 months?
  - How many people do you know who are middle school teachers?
  - ...
  - How many people do you know named Michael?
- "Know" typically defined: you know them and they know you and have you been in contact with them over the past two years

http://bit.ly/socnet204

Next class:

- Using social networks to estimate the sizes of the groups most at-risk for HIV