

Networks and hidden populations at-risk for HIV

Matthew J. Salganik

Social Network (Soc 204)
Spring 2017
Princeton University

April 17, 2017



Logistics:

- ▶ Homework due Wednesday (next Wednesday, and next, next Wednesday)

Sampling Immigrants from Their Social Networks – Innovations and Applications

Maria-Giovanna Merli, Professor Public Policy and Global Health,
Duke Sanford School of Public Policy
April 18, 2017, Noon, 300 Wallace Hall

<http://opr.princeton.edu/seminars/spring/2017>

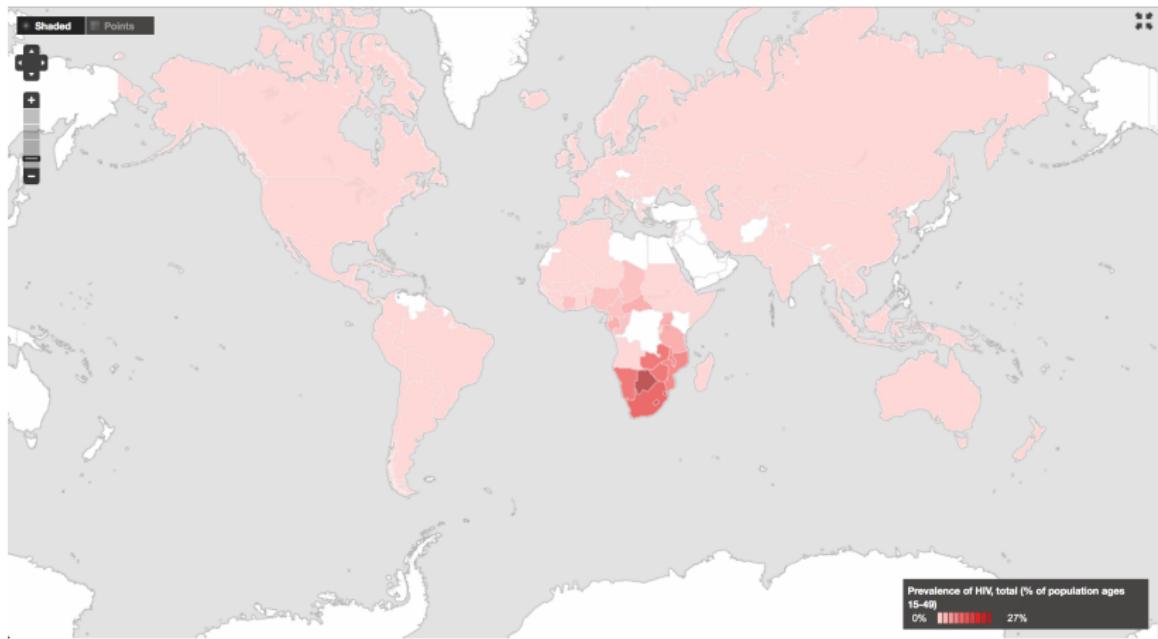
Review of last week:

- ▶ we can use surveys to collect information about personal networks of respondents
- ▶ the alters of a random sample of respondents are going to be different from a random sample of people

Vote:

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

Background



<http://data.worldbank.org/indicator/SH.DYN.AIDS.ZS/countries?display=map>

Two main stages of HIV/AIDS epidemic in a country:

- ▶ **Concentrated**: infections mostly occurring in high-risk groups (e.g., sex workers, drug injectors, men who have sex with men)
- ▶ **Generalized**: more than 1% of the general population infected

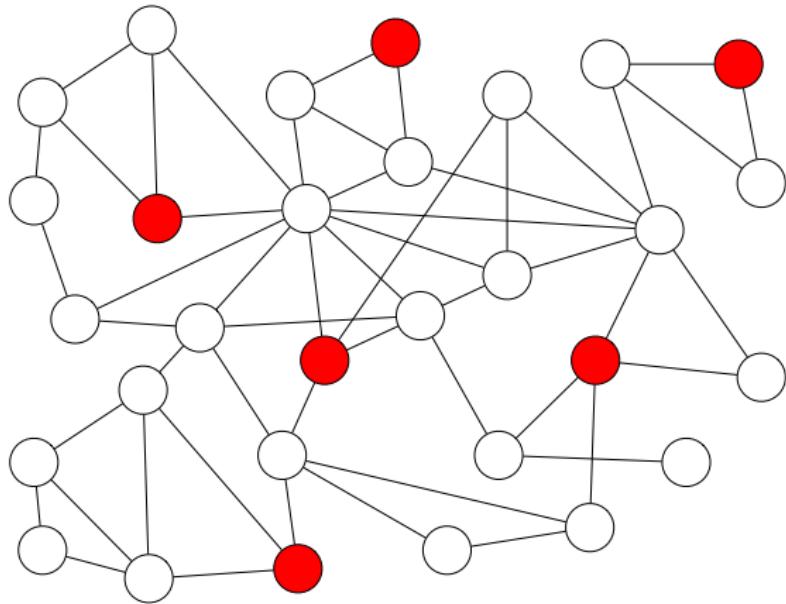
Different stages of epidemic require different research and different policy interventions

Network scale-up method

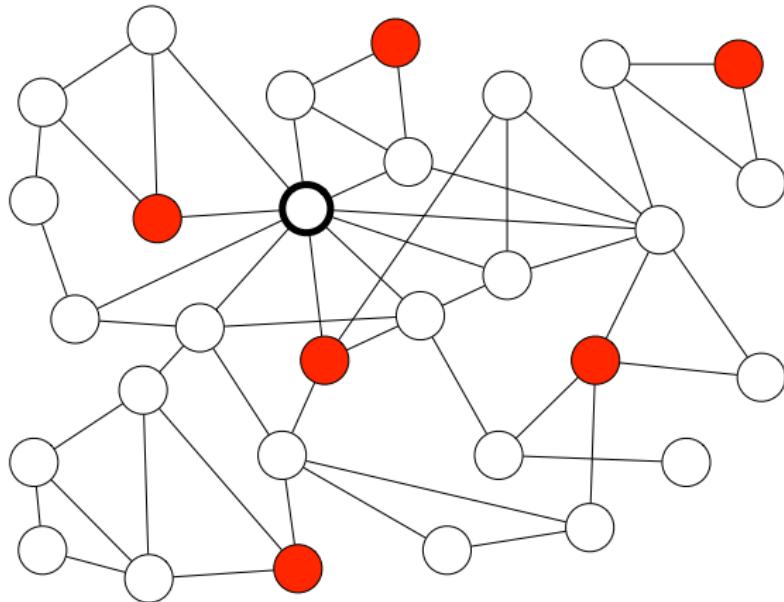


Basic insight from Bernard et al. (1989)

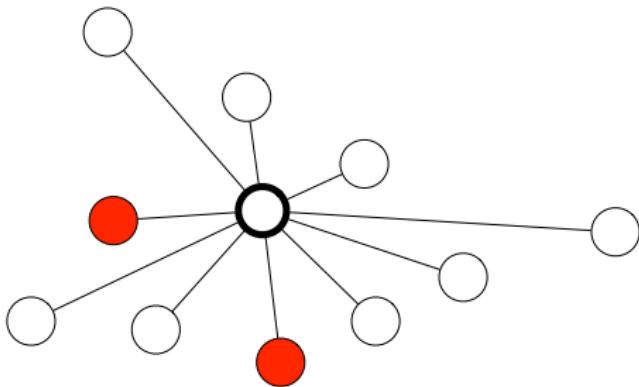
Network scale-up method



Network scale-up method



Network scale-up method



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

- ▶ 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

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

- ▶ \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)

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

See Killworth et al., (1998)

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

There are 50,000 Nsabimanas in Rwanda and 10 million people in Rwanda. You know 2 Nsabimanas. We estimate you know:

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

There are 50,000 Nsabimanas in Rwanda and 10 million people in Rwanda. You know 2 Nsabimanas. We estimate you know:

$$\hat{d}_i = \frac{2}{50,000} \times 10 \text{ million} = 400 \text{ people}$$

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

There are 50,000 Nsabimanas in Rwanda; 1,000 Priests; and 10 million people in Rwanda. You know 2 Nsabimanas and 1 priest. We estimate you know:

$$\hat{d}_i = \frac{\sum_k y_{i,k}}{\sum_k N_k} \times N$$

- ▶ \hat{d}_i : estimated number of people known by person i
- ▶ $y_{i,k}$: number of people in group k known by person i
- ▶ N_k : number of people in group k
- ▶ N : number of people in the population

There are 50,000 Nsabimanas in Rwanda; 1,000 Priests; and 10 million people in Rwanda. You know 2 Nsabimanas and 1 priest. We estimate you know:

$$\hat{d}_i = \frac{2 + 1}{50,000 + 1,000} \times 10 \text{ million} \approx 600 \text{ people}$$

Compared to direct estimation, scale-up provides:

Compared to direct estimation, scale-up provides:

- ▶ more information per respondent

Compared to direct estimation, scale-up provides:

- ▶ more information per respondent
- ▶ respondents don't have to report about themselves

Compared to direct estimation, scale-up provides:

- ▶ more information per respondent
- ▶ respondents don't have to report about themselves
- ▶ opportunity to learn about people you didn't interview

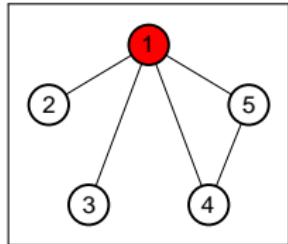
Compared to direct estimation, scale-up provides:

- ▶ more information per respondent
- ▶ respondents don't have to report about themselves
- ▶ opportunity to learn about people you didn't interview
- ▶ partially self-validating

Unfortunately, scale-up is also more complicated. But, solving problems to make scale-up work provides general knowledge about social networks around the world.

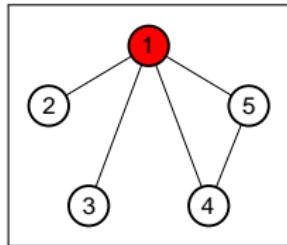
How might the friendship paradox impact estimates?

How might the friendship paradox impact estimates?
Set of egos can be different from set of alters.

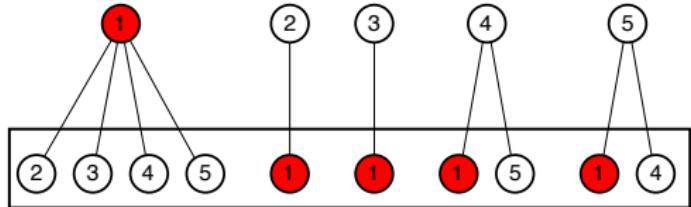


$$p = 0.2$$

How might the friendship paradox impact estimates?
Set of egos can be different from set of alters.

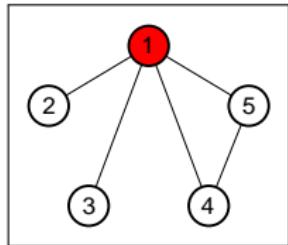


$$p = 0.2$$

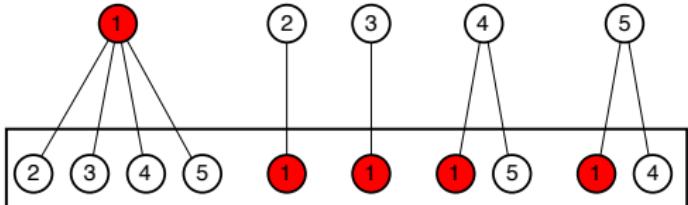


$$p_{alter} = 0.4$$

How might the friendship paradox impact estimates?
Set of egos can be different from set of alters.



$$p = 0.2$$

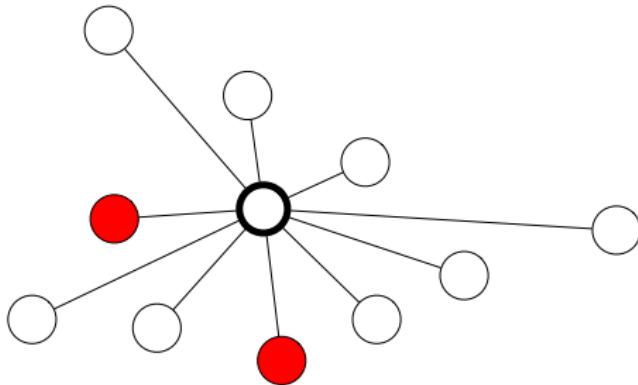


$$p_{alter} = 0.4$$

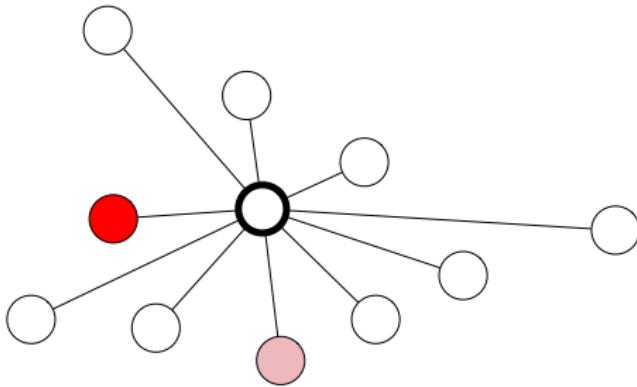
$$p_{alter} = p \times \frac{\text{avg. degree (hidden pop.)}}{\text{avg. degree (general pop.)}} = p\delta$$

Estimates will be biased by a factor of δ ("degree ratio")

How might imperfect knowledge impact scale-up estimates?
Ego is not aware of everything about all of their alters.



How might imperfect knowledge impact scale-up estimates?
Ego is not aware of everything about all of their alters.



Estimates will be biased by a factor of τ ("information transmission rate")

More about this on Wednesday!

Generalized scale-up estimator:

$$\hat{p} = \frac{\sum_i y_i}{\sum_i \hat{d}_i} \cdot \left(\frac{1}{\hat{\delta}}\right) \cdot \left(\frac{1}{\hat{\tau}}\right)$$

GENERALIZING THE NETWORK SCALE-UP METHOD: A NEW ESTIMATOR FOR THE SIZE OF HIDDEN POPULATIONS

*Dennis M. Feehan**
Matthew J. Salganik†

<http://dx.doi.org/10.1177/0081175016665425>

Network scale-up study in Rwanda

Study in Rwanda was designed to estimate the number of

- ▶ men who have sex with men
- ▶ female sex workers
- ▶ clients of female sex workers
- ▶ injection drug users

AND

to produce generalizable knowledge about the scale-up method

Quest for
fundamental
understanding?

Consideration of use?

	Yes	No
Yes	Pure basic research (Bohr)	Use-inspired basic research (Pasteur)
No		Pure applied research (Edison)

- ▶ Why was UNAIDS excited about this study?

- ▶ Why was UNAIDS excited about this study?
- ▶ Why was the Rwandan National AIDS program excited about this study?

- ▶ Why was UNAIDS excited about this study?
- ▶ Why was the Rwandan National AIDS program excited about this study?
- ▶ Why was I excited about this study?

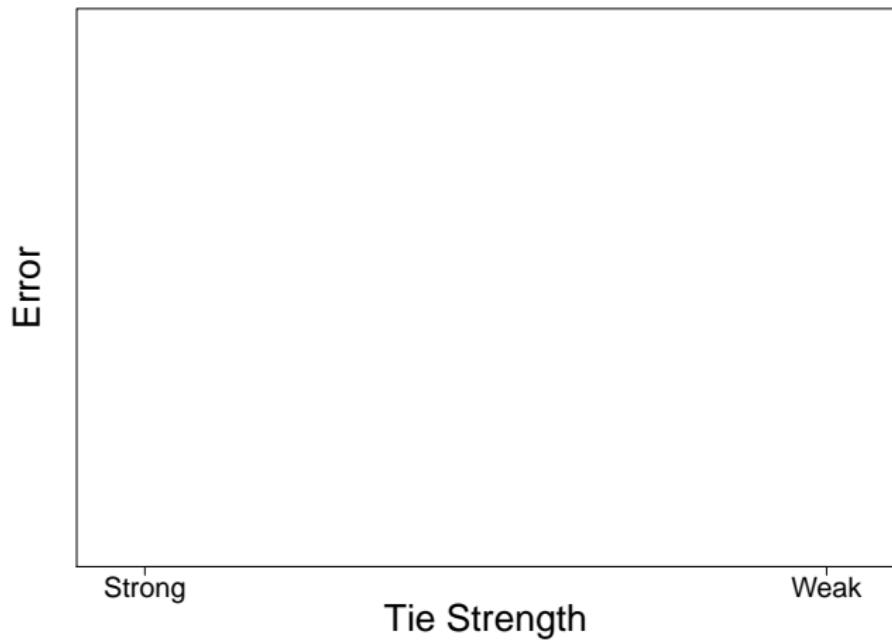


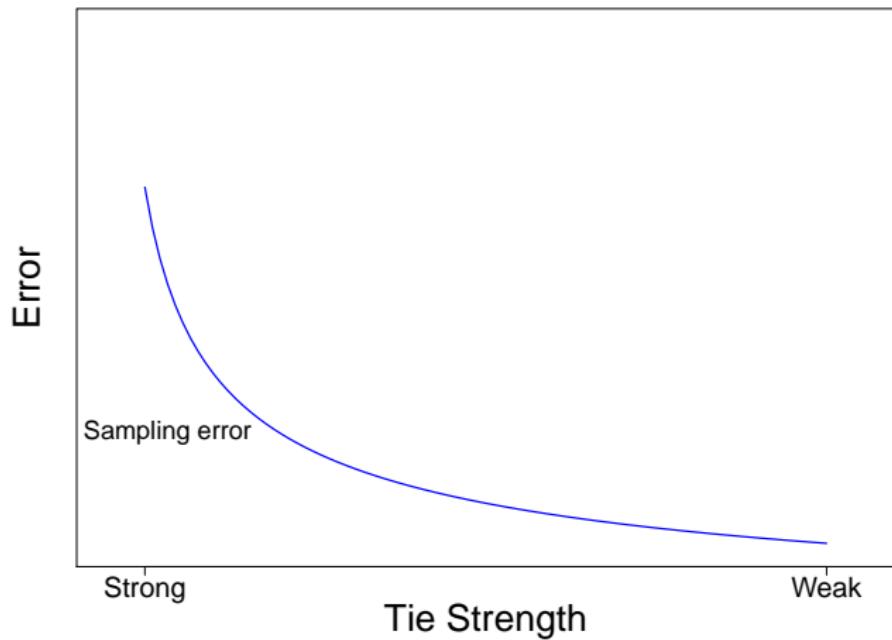


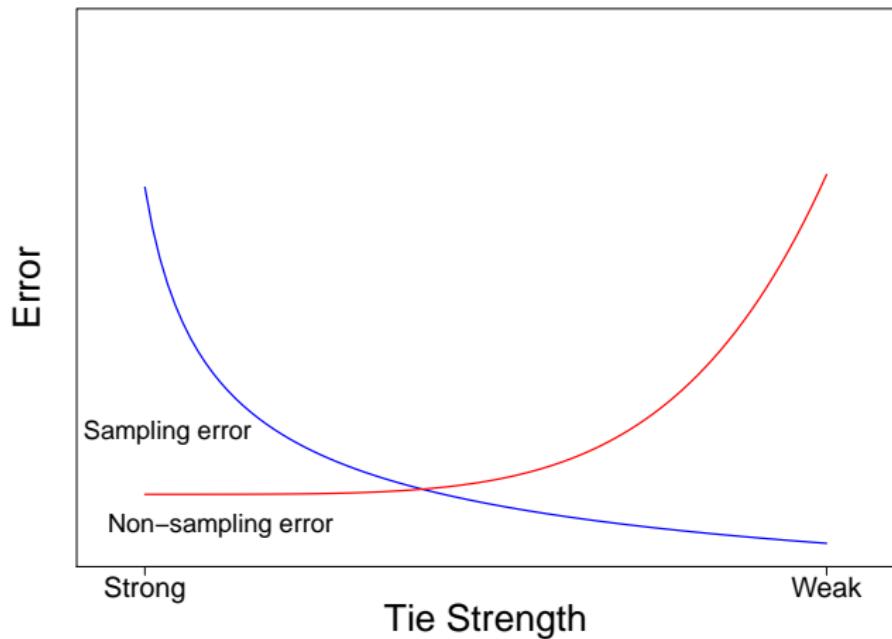


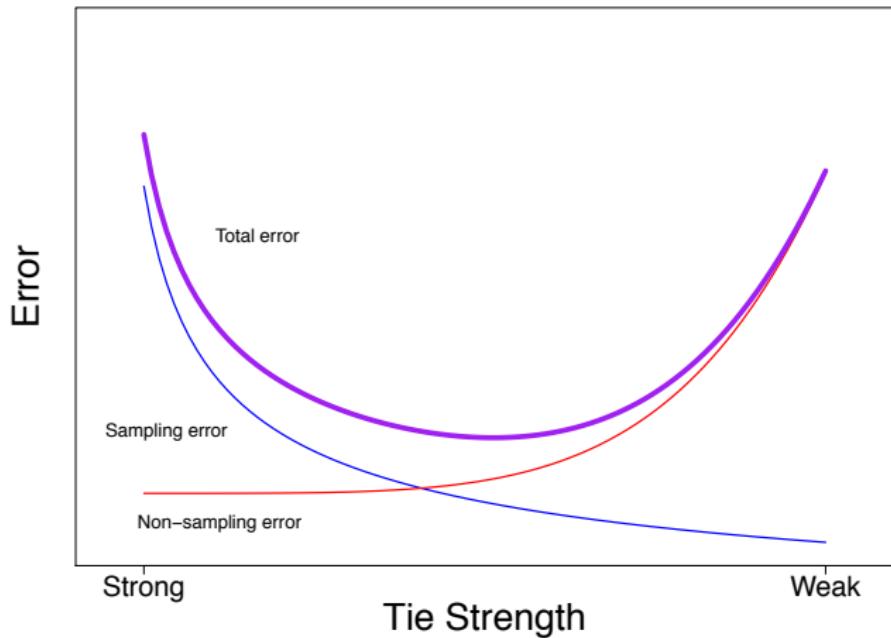


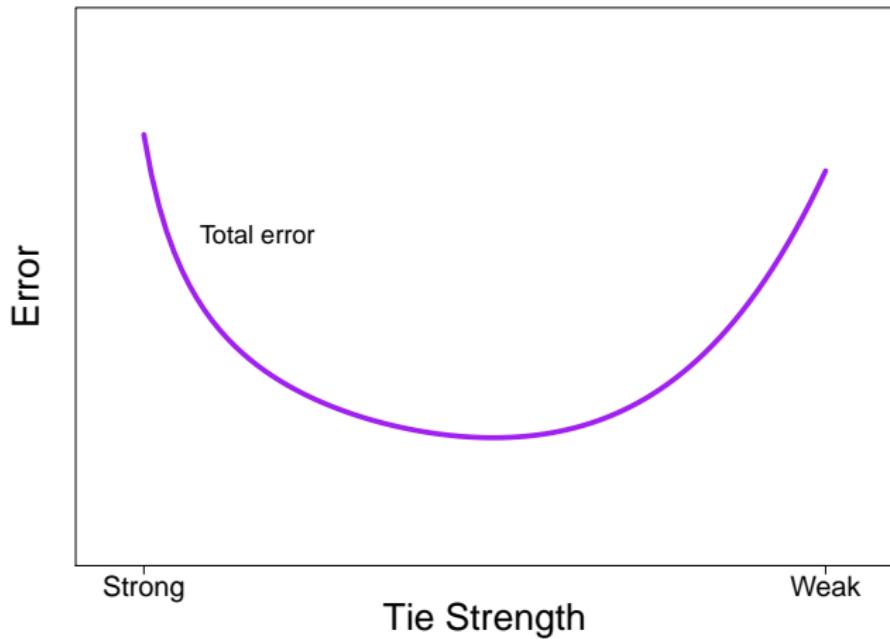












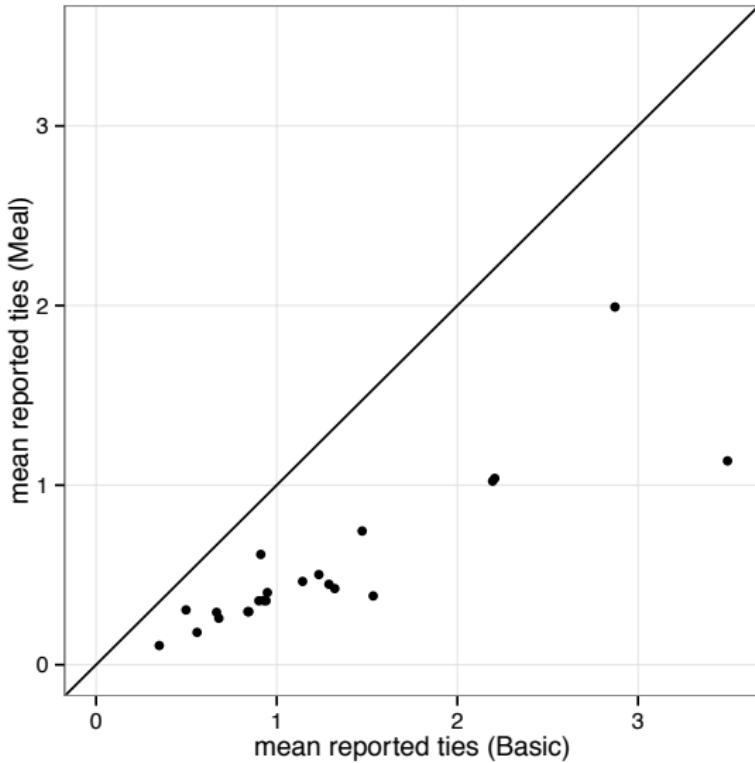
Basic definition ($n = 2,500$)

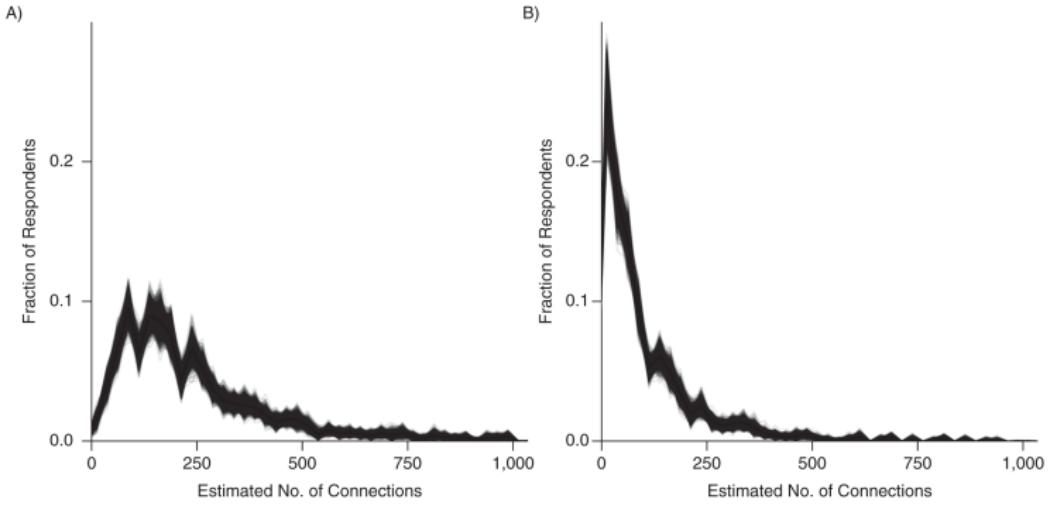
- ▶ people you know by sight and name and who also know you by sight and name
- ▶ people you have **had some contact with** in the past 12 months
- ▶ people of all ages who live in Rwanda

Meal definition ($n = 2,500$)

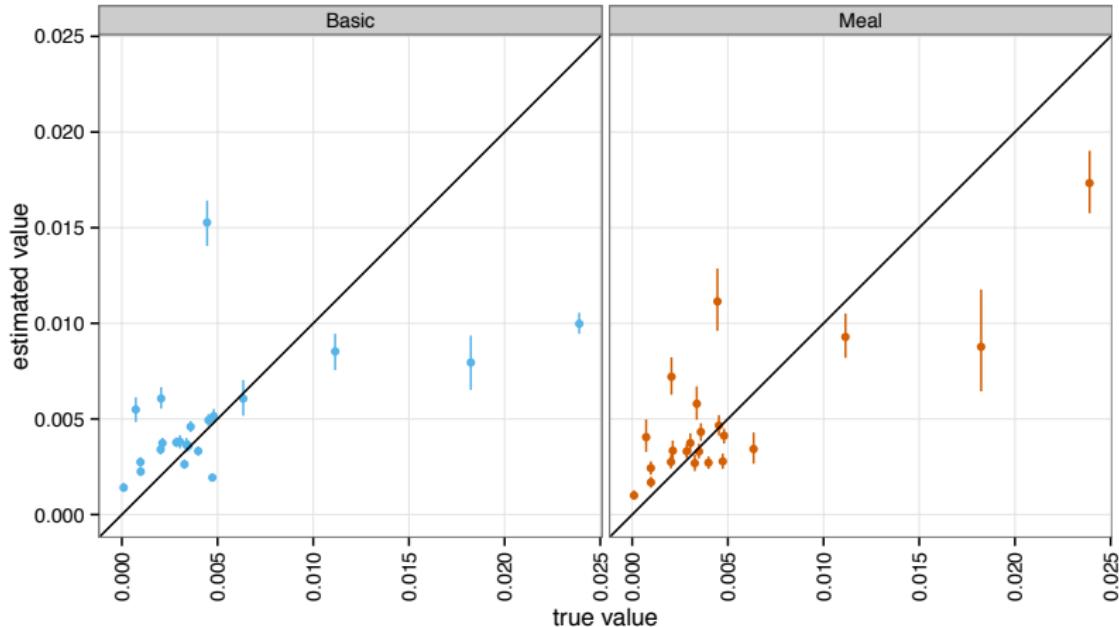
- ▶ people you know by sight and name and who also know you by sight and name
- ▶ people you have **shared a meal or drink with** in the past 12 months
- ▶ people of all ages who live in Rwanda

Priests	Twahirwa
Nurses or Doctors	Mukandekezi
Male Community Health Worker	Nyiraneza
Widowers	Ndayambaje
Teachers	Murekatete
Divorced Men	Nsengimana
Incarcerated people	Mukandayisenga
Women who smoke	Ndagijimana
Muslim	Bizimana
Women who gave birth in the last 12 mo.	Nyirahabimana
	Nsabimana
	Mukamana

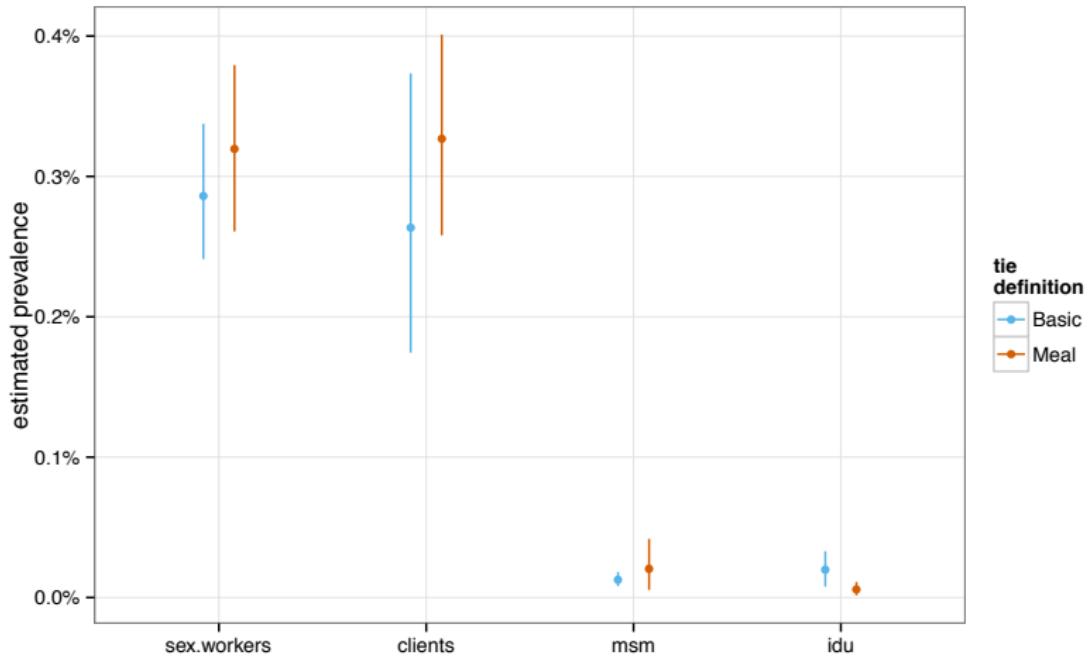




Is either degree distribution power-law?



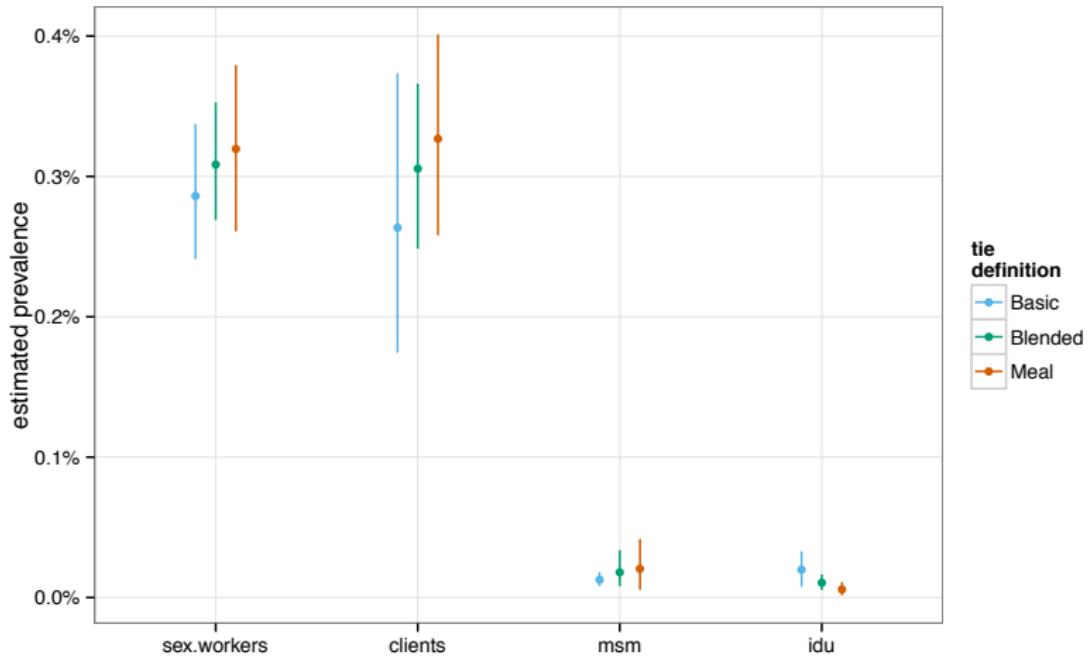
Meal definition has lower error (RMSE, MAE, MRE)



$$\hat{N}_H = w \cdot \hat{N}_{H[meal]} + (1 - w) \cdot \hat{N}_{H[basic]}$$

$$\hat{N}_H = w \cdot \hat{N}_{H[meal]} + (1 - w) \cdot \hat{N}_{H[basic]}$$

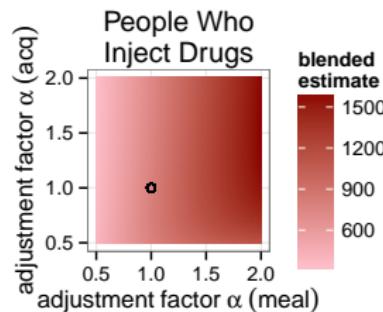
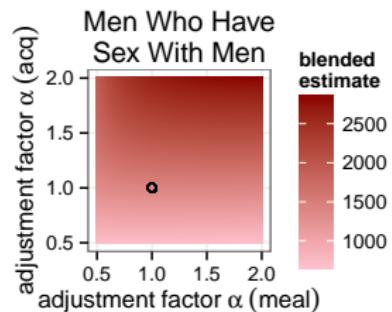
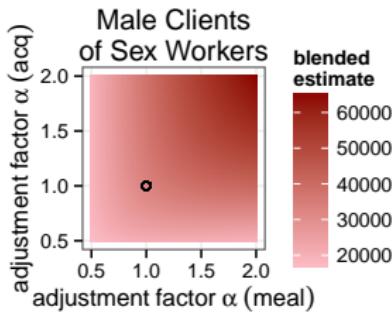
$$w = \frac{\hat{\sigma}_{basic}^2}{\hat{\sigma}_{basic}^2 + \hat{\sigma}_{meal}^2}$$

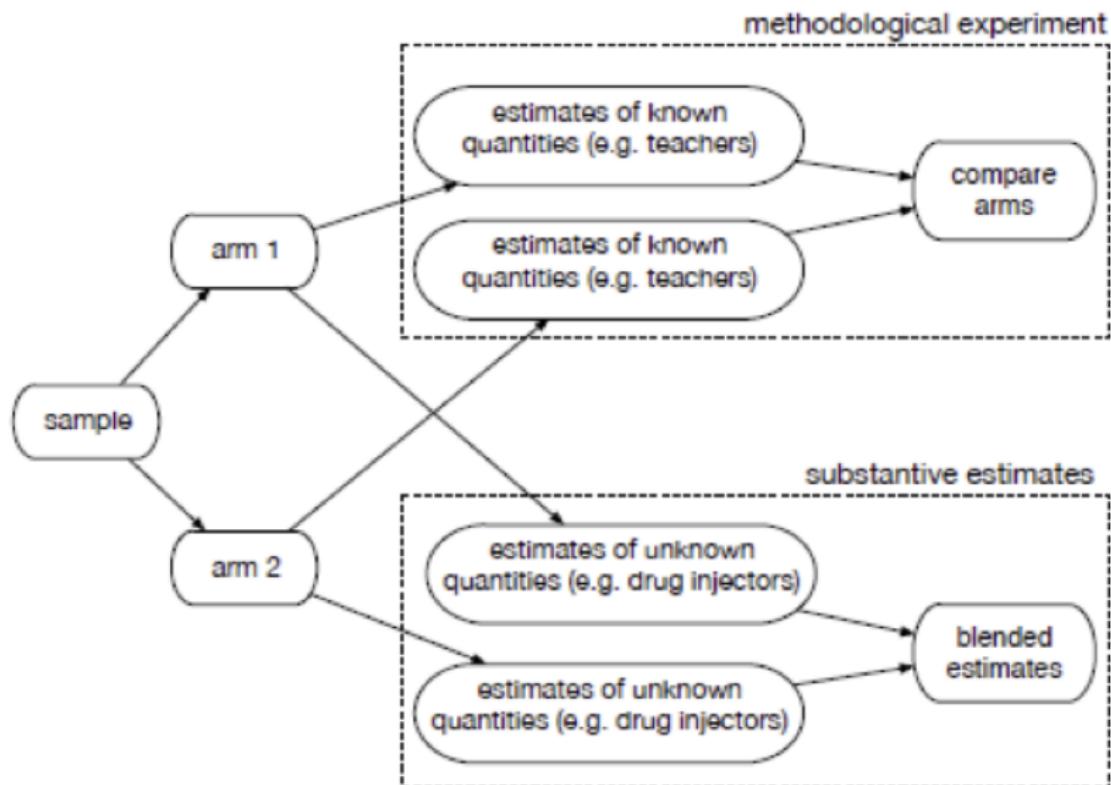


$$N_H = \alpha \hat{N}_H$$

where

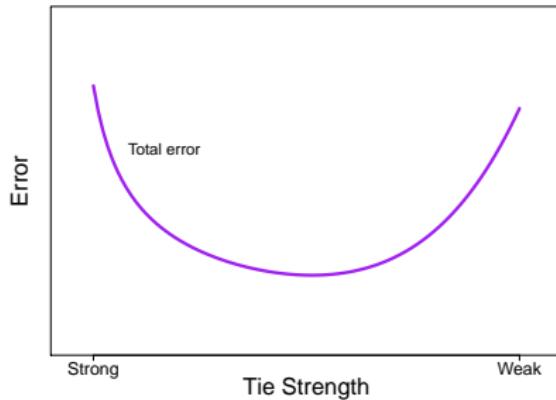
$$\alpha = \underbrace{\left(\frac{\eta_F}{\tau_F} \right)}_{\text{reporting distortions}} \times \underbrace{\left(\frac{1}{\phi_F \delta_F} \right)}_{\text{structural distortions}}$$



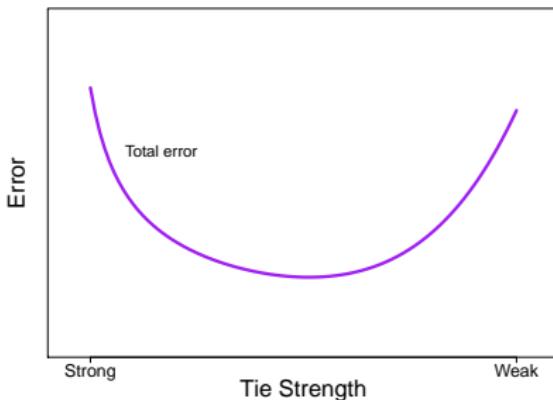


Next steps for network scale-up method

With only two arms, we cannot demonstrate U-shaped relationship!



With only two arms, we cannot demonstrate U-shaped relationship!



So, we are going to do a survey experiment with 4 arms:

1. shared a meal with yesterday
2. shared a meal with in the past seven days
3. shared a meal with this semester
4. shared a meal with this academic year

<http://bit.ly/socnet204>

Next class:

- ▶ Who knows what about who?