Discussion of

# Estimating Social Networks Models with Missing Links

Lewbel, Qu and Tang (2023)

Yong Cai

March 31, 2023

# Outline

# Summary

# Setting

- Peer effects regression:

$$y = \lambda G y + X\beta + \varepsilon$$

  where $G$ is adjacency matrix of the network

- Many results on identification and estimation when $G$ is perfectly observed

- Less is known when $G$ is unobserved or observed with error

# Network Data Often Unobserved or Observed with Error

- Network data is high dimensional and thus costly to collect
- To limit data collection, surveyors may ask respondees to list $X$ friends
- Respondees may not be able to recall all connections
- Relationships are intensities, hard to quantify and elicit

# This Paper

Peer effects regression when network data is missing at random

1. Shows that augmentation bias arises

2. Provides 2SLS-based solution when multiple networks are observed

# Augmentation Bias

- Let $H$ be observed adj. matrix with $p$ proportion of links randomly missing
- Using $H$ as plug-in for $G$ leads to augmentation bias:

$$y = \left(\frac{\lambda}{1-p}\right) Hy + X\beta + v \quad , \quad E[v|X,G] = 0$$

- In OLS: attenuation bias with mean zero white noise measurement error
- Missingness has negative mean
- Intuition: an individual is affected by 5 friends but we misattribute to 3

# 2SLS with Multiple Networks

- $Gy$ is endogenous; use $GX$ or $G^2X$ as "friends-of-friends" instruments
- Not possible to use $HX$ or $H^2X$ as instruments
- With multiple independent networks, $H^{(2)}X$ can instrument for $H^{(1)}y$ to estimate $\frac{\lambda}{1-p}$
- Estimate $p$ by looking at how many links observed in $H^{(2)}$ are missing in $H^{(1)}$

# Discussion

# What if adjacency matrix is row-normalized?

- Adjacency matrix is often row-normalized:

$$y_i = \lambda \left( \frac{1}{G_i} \sum_{j=1}^{n} y_j G_{ij} \right) + X_i \beta + \varepsilon_i \quad , \quad G_i = \sum_{j=1} G_{ij}$$

- Denominator is also changing $\Rightarrow$ no/attenuation bias?

## When do we observe multiple copies of the same network?

- Multiple networks may be collected, but they seem different
  - Not clear that network of loans is network of friendships with more missingness
  - Depending on the outcome, not clear if either is exactly the network of interest
- Asymmetric networks seem to reflect asymmetric relations
  - If $i$ visits $j$ but $j$ does not visit $i$, maybe $i$ is influenced by $j$ but not vice versa
  - Networks data often symmetrized in practice, but maybe asymmetry might be important (Comola and Fafchamps, 2014; Auerbach, 2019; Gao, Li, and Xu, 2022)
- Will matrix completion under a low-rank assumption work instead?

# Is missingness random in practice?

- Stronger links may be more likely to be reported (Griffith, 2022)
- Agents may have incentive to misreport links (Comola and Fafchamps, 2017)
- What type of non-random missingness can be accommodated?
- Lewbel et al. (2023, JPE) assumes that network is unobserved. Can these estimates be used for a test on missingness?

# References

Auerbach, E. (2019). Testing for differences in stochastic network structure. *arXiv preprint arXiv:1903.11117*.

Comola, M. and M. Fafchamps (2014). Testing unilateral and bilateral link formation. *The Economic Journal 124*(579), 954–976.

Comola, M. and M. Fafchamps (2017). The missing transfers: Estimating misreporting in dyadic data. *Economic Development and Cultural Change 65*(3), 549–582.

Gao, W. Y., M. Li, and S. Xu (2022). Logical differencing in dyadic network formation models with nontransferable utilities. *Journal of Econometrics*.

Griffith, A. (2022). Name your friends, but only five? the importance of censoring in peer effects estimates using social network data. *Journal of Labor Economics 40*(4), 779–805.

Lewbel, A., X. Qu, and X. Tang (2023). Social networks with unobserved links. *Journal of Political Economy 131*(4), 000–000.