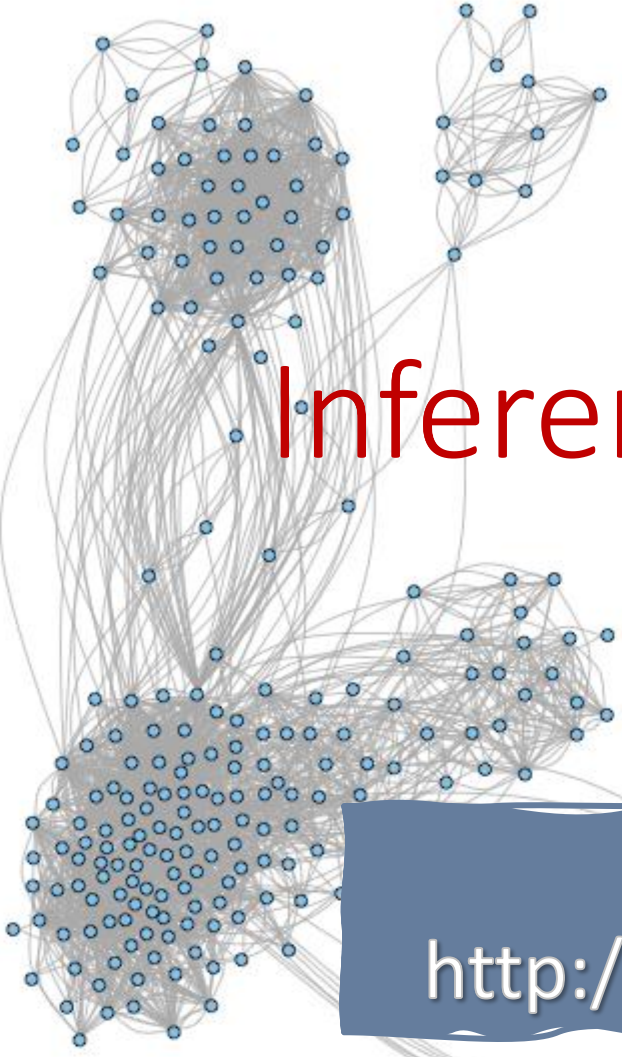
Cornell University – Department of Sociology



# A short course on
# Inferential social network analysis

Benjamin Rosche and Wicia Fang

Get the materials:
http://benrosche.com/teaching/isna-workshop/

# Content

1. Introduction: Social networks and regression modeling
2. Network regression models
3. R workshop
4. Conclusions

# 1 Social networks and regression modeling

- Two types of inferential social network analysis (SNA):
  1. Models of networks – network measures as dependent variables
  2. Models of network effects – network measures as explanatory variables **today**
- In social and political systems, interactions between actors are an integral part of the process of interest.
- SNA extends of the conventional regression model by:
  1. Accounting for *interdependence of observations*
  2. Pose and test new questions and theories
- Example: Do students' SES predict their academic achievement?
  - $GPA_i = \beta_0 + \beta_1 SES_i + e_i$

# 1 Networks effects

## Three ways in which network embeddedness can affect individual outcomes (An, Beauville, Rosche 2022)

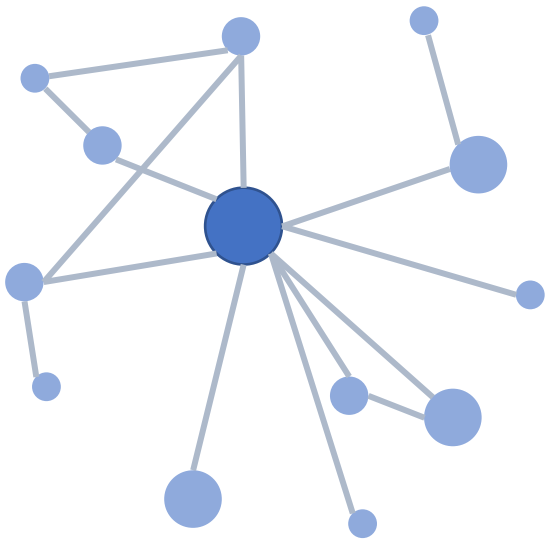
- Peer effects: $GPA_i = \beta GPA_j$ **today**

# 1 Networks effects

Three ways in which network embeddedness can affect individual outcomes (An, Beauville, Rosche 2022)

- Peer effects: $GPA_i = \beta GPA_j$ **today**

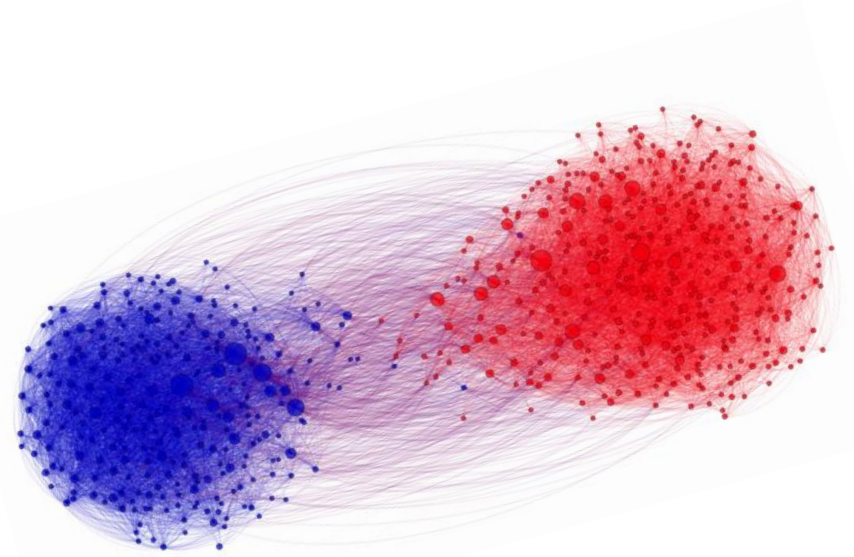
- Positional effects: $GPA_i = \beta Centrality_i$



# 1 Networks effects

## Three ways in which network embeddedness can affect individual outcomes (An, Beauville, Rosche 2022)

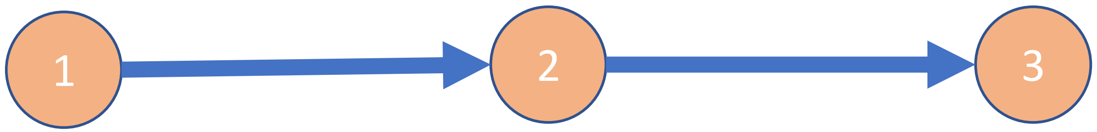
- Peer effects: $GPA_i = \beta GPA_j$ **today**
- Positional effects: $GPA_i = \beta Centrality_i$
- Structural effects: $GPA_i = \beta Polarization_{N(i)}$



# 1 Networks effects

## Three types of peer effects

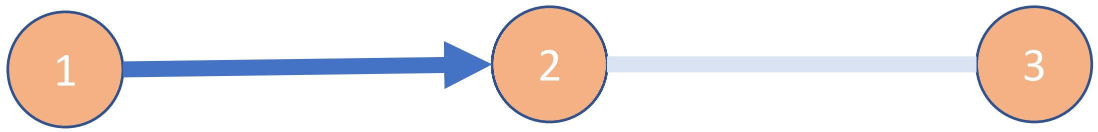
- Endogenous peer effect (global impact): $GPA_i = \beta GPA_j$ **today**



# 1 Networks effects

## Three types of peer effects

- Endogenous peer effect (global impact): $GPA_i = \beta GPA_j$ **today**
- Exogenous peer effect (local impact): $GPA_i = \beta SES_j$ **today**



# 1 Networks effects

## Three types of peer effects

- Endogenous peer effect (global impact): $GPA_i = \beta GPA_j$ **today**
- Exogenous peer effect (local impact): $GPA_i = \beta SES_j$ **today**
- Disturbance peer effect: $u_i = \beta u_j + \varepsilon_i$

# 1 Networks effects

## Different peer groups

- best-friend effect: $GPA_i = \beta GPA_j$
- peer group effect: $GPA_i = \beta \overline{GPA}_{-i}$
- peer network effect: $GPA_i = \beta w_{ij}^{*} GPA_j$ **today**

where $w_{ij}^{*} = w_{ij} / \sum w_{ij}$

adjacency matrix

$$\boldsymbol{w} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

# 2 Network regression models

## Network regression models can be differentiated by

- whether or not they consider that the network might have formed endogenously
- whether they are cross-sectional or panel

## Today: focus on models that assume **w** to be exogenous

- Linear-in-means model (peer group effect)
- Spatial autoregressive models (peer network effect) **today**

# 2 Varieties of spatial regression models

- Spatial autoregressive model: $\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ **today**
- Spatial lagged-X model: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$ **today**
- Spatial error model: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$
- Spatial Durbin model: $\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$
- Spatial autoregressive combined model: $\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$
- Spatial Durbin error model: $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$
- General spatial model*: $\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{X}\boldsymbol{\beta} + \lambda \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon}$

* = weakly identifiable

# Spatial lagged-X model

## Cross-sectional model

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{WX\theta} + \boldsymbol{\varepsilon}$$

$\boldsymbol{\theta}$ = exogenous peer effects

## Panel model

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}_t\boldsymbol{\theta} + \boldsymbol{\mu} + \xi_t + \boldsymbol{\varepsilon}_t$$

$\boldsymbol{\mu}$ = individual-specific effects, $\xi_t$ = time-specific effect

## Dynamic panel model

$$\boldsymbol{y}_t = \boldsymbol{X}_t\boldsymbol{\beta} + \boldsymbol{W}_t\boldsymbol{X}_t\boldsymbol{\theta} + \boldsymbol{W}_{t-1}\boldsymbol{X}_{t-1}\boldsymbol{\eta} + \boldsymbol{\mu} + \xi_t + \boldsymbol{\varepsilon}_t$$

$\eta$ = lagged exogenous peer effects

# Spatial autoregressive model

## Cross-sectional model

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$\rho$ = endogenous peer effects

## Panel model

$$\boldsymbol{y}_t = \rho \boldsymbol{W}_t \boldsymbol{y}_t + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\mu} + \xi_t + \boldsymbol{\varepsilon}_t$$

$\boldsymbol{\mu}$ = space-specific effects, $\xi_t$ = time-specific effect

## Dynamic panel model

$$\boldsymbol{y}_t = \rho \boldsymbol{W}_t \boldsymbol{y}_t + \tau \boldsymbol{W}_{t-1} \boldsymbol{y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\mu} + \xi_t + \boldsymbol{\varepsilon}_t$$

$\tau$ = lagged exogenous peer effects

# R workshop

http://benrosche.com/teaching/isna-workshop/R-tutorial.html

# Conclusion: When are individual effects biased?

- The effect of individual features (β) will not be biased in the presence of peer effects if networks are random
- Endogenous network formation does not bias β in the presence of an endogenous peer effect (but: total effect is underestimated)
- However, β will be biased if the individual feature is part of the selection process and also influences peers (exogenous peer effect)

  e.g., the role of socioeconomic status on student achievement
- These results hold for cross-sectional and panel data using a RE estimator. The FE estimator cannot estimate time-constant individual effects

* Note that these results hold for randomly distributed features. Individual features that are themselves affected by peer effects (i.e., $\boldsymbol{X} = \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta}$) will be affected more by endogenous network formation.

# Conclusion: When are peer effects biased?

Cross-sectional data

- If networks are random, exogenous peer effects are estimated correctly even if there are omitted variables
- If networks are random, the endogenous peer effect is biased if there are omitted variables
- Both exogenous and endogenous peer effects are biased if networks formed endogenously

# Conclusion: When are peer effects biased?

## Panel data

- If networks evolve over time, we can estimate time-constant exogenous peer effects using a FE estimator
- If networks are random, the FE model recovers the correct exogenous peer effects even if other exogenous peer effects are omitted
- The FE model also recovers correct exogenous peer effects in the presence of time-constant selection effects if all relevant exogenous peer effects are included in the model
- The FE model recovers the correct endogenous peer effect as long as the selection process is time-constant (!)

# Thank you for your attention!

## Helpful reviews

Hsieh, Lin, Patacchini (2020), Bramoulle, Djebbari, Fortin (2020), Ruttenauer (2022), An, Beauville, Rosche (2022)

## Advanced parametric models that account for the endogeneity of w

SAOM (Snijders 2011), Han, Hsieh, Ko (2021), Heckman-style two-stage approaches (Goldsmith-Pinkham & Imbens2013, Arduini et al. 2015, Hsieh & Lee 2016)