Peer Effects and Social Networks in Education

Antoni Calvó-Armengol Eleonora Patacchini Yves Zenou

Group Members

3+2巧克力味夹心饼

CONTENTS

0. Introduction

1. Model

2. Data and Empirical Analysis

3. Empirical Results and Peer Effects

4. Extension Ⅰ： Alternative Formulations

5. Extension Ⅱ： Directed Networks

Part 0

Introduction

0.1 About the Authors

Antoni Calvó-Armengol

Eleonora Patacchini

Yves Zenou

0.2 The Novelty of this Paper

1. Stress the role of the structure of social networks in explaining individual behaviour, from a conceptual point of view

Social networks → Individual behaviour

2. Build a theoretical model of peer effects that envisions group influence as an equilibrium outcome, from a more operational point of view

Nash equilibrium — Katz-Bonacich centrality

3. Conduct a direct empirical test of the model using a detailed dataset on the friendship networks, Add Health, with particular attention to the relevant econometric problems

Part 1

Model

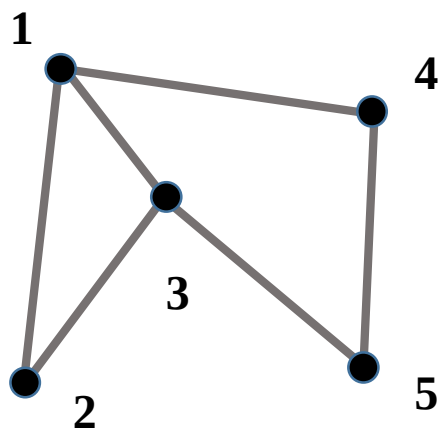
1.1 Networks

$N = \{1, \ldots, n\}$ is a finite set of agents

network $\mathbf{g}$ $\qquad \mathbf{G} = [g_{ij}]$

$$g_{ij} = \begin{cases} 1 \text{ if } i \text{ and } j \text{ are direct friends} \\ 0, \text{ otherwise} \end{cases}$$

$$g_{ii} = 0 \qquad g_{ij} = g_{ji}$$



$$\begin{bmatrix} 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

1.2 Preference

$$u_i(\mathbf{y}^0, \mathbf{z}; \mathbf{g}) = \boxed{\theta_i\, y_i^0 - \frac{1}{2}\left(y_i^0\right)^2} + \boxed{\mu g_i z_i - \frac{1}{2} z_i^2 + \phi \sum_{j=1}^{n} g_{ij} z_i z_j} \qquad \phi > 0,\ \mu > 0$$

$$\theta_i(\mathbf{x}) = \sum_{m=1}^{M} \beta_m x_i^m + \frac{1}{g_i} \sum_{m=1}^{M} \sum_{j=1}^{n} \gamma_m g_{ij}\, x_j^m$$

$g_i = \sum_{j=1}^{n} g_{ij}$ is the number of direct links of individual i

x_i^m is a set of M variables accounting for observable differences

$$\frac{\partial^2 u_i(\mathbf{y}^0, \mathbf{z}; \mathbf{g})}{\partial z_i \partial z_j} = \phi g_{ij} \geq 0 \qquad i \neq j$$

1.3 Katz-Bonacich Centrality

$$\mathbf{b}(\mathbf{g}, \phi) = \phi\mathbf{G}\mathbf{1} + \phi^2\mathbf{G}^2\mathbf{1} + \phi^3\mathbf{G}^3\mathbf{1} + \cdots = \sum_{k=0}^{+\infty} \phi^k \mathbf{G}^k \cdot (\phi\mathbf{G}\mathbf{1})$$

$$\mathbf{b}(\mathbf{g}, \phi) = (\mathbf{I} - \phi\mathbf{G})^{-1} \cdot (\phi\mathbf{G}\mathbf{1})$$

$$(\mathbf{G}^2)_{ij} = \sum_{s=1}^{n} g_{is} g_{sj}$$

$\mathbf{G}^k$ gives the number of paths of length k in $\mathbf{g}$ between i and j

$\mathbf{G}^k\mathbf{1}$ give the total number of paths of length k that emanate from the corresponding network node

$\phi\omega(\mathbf{g}) < 1$ Denote by $\omega(\mathbf{g})$ the largest eigenvalue of the adjacency matrix $\mathbf{G} = [g_{ij}]$ of the network

$$\sum_{j \neq i} \left| \frac{\partial^2 u_i}{\partial z_i \partial z_j} \right| < \left| \frac{\partial^2 u_i}{\partial z_i^2} \right|, \quad \text{for all } i = 1, \ldots, n$$

network $\mathbf{g}$ $\qquad \mathbf{G} = [g_{ij}]$

$$u_i(\mathbf{y}^0, \mathbf{z}; \mathbf{g}) = \theta_i\, y_i^0 - \frac{1}{2}\left(y_i^0\right)^2 + \mu g_i z_i - \frac{1}{2} z_i^2 + \phi \sum_{j=1}^{n} g_{ij} z_i z_j$$

$$\mathbf{b}(\mathbf{g}, \phi) = \phi\mathbf{G}\mathbf{1} + \phi^2\mathbf{G}^2\mathbf{1} + \phi^3\mathbf{G}^3\mathbf{1} + \cdots = \sum_{k=0}^{+\infty} \phi^k \mathbf{G}^k \cdot (\phi\mathbf{G}\mathbf{1})$$

$$= (\mathbf{I} - \phi\mathbf{G})^{-1} \cdot (\phi\mathbf{G}\mathbf{1}) \qquad \phi\omega(\mathbf{g}) < 1$$

1.4 Equilibrium

$$max_{y_i^0, z_i}\{ u_i(\mathbf{y}^0, \mathbf{z}; \mathbf{g}) = \theta_i\, y_i^0 - \frac{1}{2}\left(y_i^0\right)^2 + \mu g_i z_i - \frac{1}{2} z_i^2 + \phi \sum_{j=1}^{n} g_{ij} z_i z_j \}$$

$$y_i^{0*}(\mathbf{x}) = \theta_i(\mathbf{x}) = \sum_{m=1}^{M} \beta_m x_i^m + \frac{1}{g_i} \sum_{m=1}^{M} \sum_{j=1}^{n} \gamma_m g_{ij}\, x_j^m$$

$$z_i^*(\mathbf{g}) = \mu g_i + \phi \sum_{j=1}^{n} g_{ij} z_j \xrightarrow{\mathbf{z}^* = \mu \mathbf{G1} + \phi \mathbf{G z}^*} z_i^*(\mathbf{g}) = \frac{\mu}{\phi} b_i(\mathbf{g}, \phi)$$

$$y_i^*(\mathbf{x}, \mathbf{g}) = \underbrace{y_i^{0*}(\mathbf{x})}_{\text{idiosyncratic}} + \underbrace{z_i^*(\mathbf{g})}_{\text{peer effect}} = \sum_{m=1}^{M} \beta_m x_i^m + \frac{1}{g_i} \sum_{m=1}^{M} \sum_{j=1}^{n} \gamma_m g_{ij}\, x_j^m + \frac{\mu}{\phi} b_i(\mathbf{g}, \phi)$$

1.5 A more general model

A more general utility function: $u_i(\mathbf{z}) = \alpha_i z_i + \frac{1}{2}\sigma_{ii} z_i^2 + \sum_{j \neq i} \sigma_{ij} z_i z_j$ $\quad \boldsymbol{\Sigma} = [\sigma_{ij}]$

Decomposition of the interaction matrix: $\boldsymbol{\Sigma} = -\beta \mathbf{I} - \gamma \mathbf{J} + \lambda \mathbf{G}$

$$\mathbf{I} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \qquad \mathbf{J} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \qquad \mathbf{G} = [g_{ij}]$$

$\mathbf{u}$-weighed Katz-Bonacich centrality: $\mathbf{w_u}(\mathbf{g}, a) = \left(\mathbf{I} - a\mathbf{G}^{-1}\right)\mathbf{u} = \sum_{p=0}^{+\infty} a^p \mathbf{G}^p \mathbf{u}$

Equilibrium: $\boldsymbol{\alpha} = \alpha \mathbf{1} \quad \mathbf{z}^* = \frac{\alpha}{\beta + \gamma w_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)} \mathbf{w_1}(\mathbf{g}, \lambda/\beta)$

Part 2

Data and Empirical Analysis

2.1 Database

The Add Health Database

- 90,118 students (grades 7–12) ➡ 20,745
- From 130 schools
- in years 1994–1995
- questionnaire

Define Friendship Networks

assume that friendship relationships are reciprocal

Sample

90,118 → 20,745 → 11,964

199 Networks

2.2 Descriptive evidence

REVIEW OF ECONOMIC STUDIES

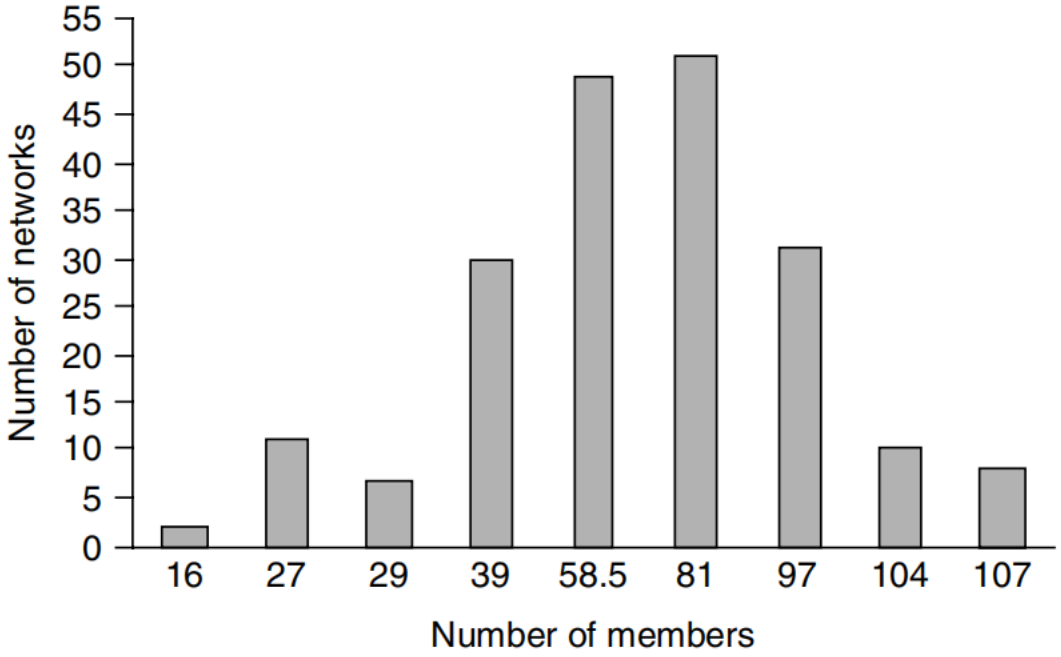


FIGURE 1

The empirical distribution of adolescent networks

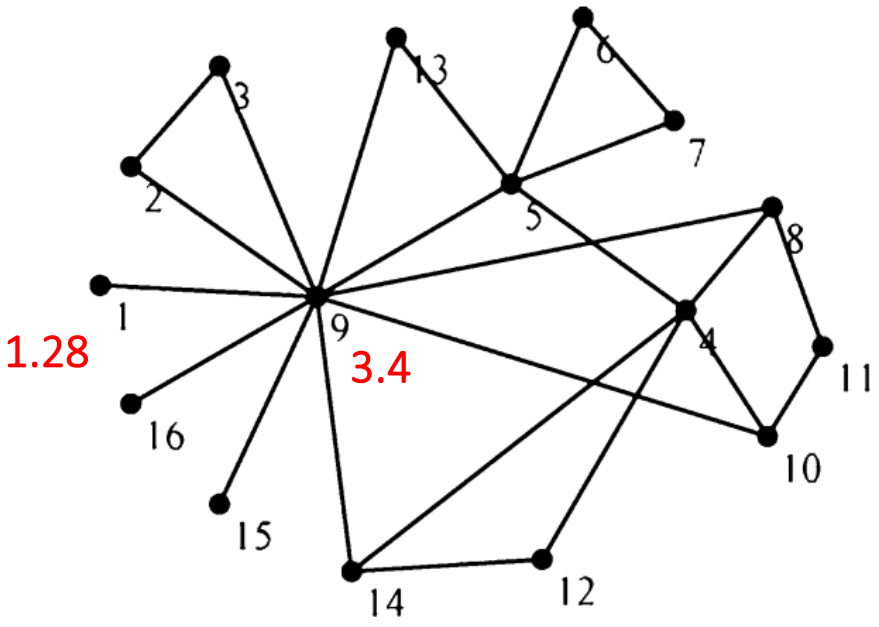


FIGURE 2

Smallest network of adolescents (n = 16)

2.3 Empirical Strategy

Proposition 1. *Suppose that $\phi\omega(\mathbf{g}) < 1$. Then, the individual equilibrium outcome is uniquely defined and given by:*

$$y_i^*(\mathbf{x}, \mathbf{g}) = \underbrace{y_i^{0*}(\mathbf{x})}_{\text{idiosyncratic}} + \underbrace{z_i^*(\mathbf{g})}_{\text{peer effect}}$$

$$y_i^{0*}(\mathbf{x}) = \theta_i(\mathbf{x}) = \sum_{m=1}^{M} \beta_m x_i^m + \frac{1}{g_i} \sum_{m=1}^{M} \sum_{j=1}^{n} \gamma_m g_{ij} x_j^m \qquad z_i^*(\mathbf{g}) = \mu g_i + \phi \sum_{j=1}^{n} g_{ij} z_j$$

Traditional Regression Model

$$y_{i,\kappa} = \boxed{\sum_{m=1}^{M} \beta_m x_{i,\kappa}^m + \frac{1}{g_{i,\kappa}} \sum_{m=1}^{M} \sum_{i=1}^{n_\kappa} \gamma_m g_{ij,\kappa} x_{j,\kappa}^m} + \eta_\kappa + \varepsilon_{i,\kappa},$$

Spatial Error Model

$$\varepsilon_{i,\kappa} = \mu g_{i,\kappa} + \phi \sum_{j=1}^{n_\kappa} g_{ij,\kappa} \varepsilon_{j,\kappa} + \upsilon_{i,\kappa}, i = 1, \ldots, n; \kappa = 1, \ldots, K,$$

2.4 Identification of Peer Effects

1 Endogenous Sorting

Network-fixed effects:

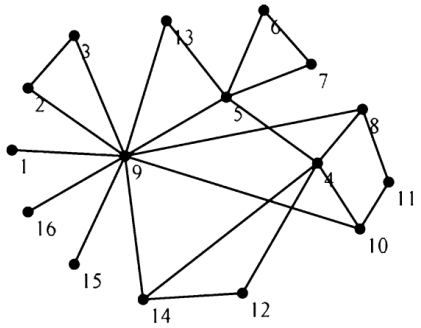
$$y_{i,\kappa} = \sum_{m=1}^{M} \beta_m x_{i,\kappa}^m + \frac{1}{g_{i,\kappa}} \sum_{m=1}^{M} \sum_{j=1}^{n_\kappa} \gamma_m g_{ij,\kappa} x_{j,\kappa}^m + \boxed{\eta_\kappa} + \varepsilon_{i,\kappa}$$

2 Reflection Problem (Manski, 1993)

Individual peer groups variation:

Proposition 2. *Suppose that $\phi\omega(\mathbf{g}) < 1$ and $\mu \neq 0$. Peer effects are identified if and only if $g_i^{[2]}/g_i \neq g_j^{[2]}/g_j$ for at least two agents i and j.*

$g_i^{[2]}/g_i$: *the average connectivity of agent i's direct contacts.*



3 Specific Control

Indicators:

- self-esteem
- physical development
- motivation in education
- parental care
- school-fixed effects

2.4 Estimate Strategy

Cross-section Regression

- Obtain 199 estimated values of φ and μ: $\widehat{\phi_\kappa}$, $\widehat{\mu_\kappa}$

Filtrate

- Calculate each network gκ's largest eigenvalue ω(gκ) and check which network does not satisfy the condition ϕκ < 1/ω(gκ), then removed.
- 181 networks remain.

Calculate

- Average estimate of ϕ and μ: $\overline{\mu_k}$, $\overline{\varphi_k}$
- $b_i(g, \Phi)$

Part 3

Empirical Results and Peer Effects

3.1 Empirical Results

$$y_{i,\kappa} = \sum_{m=1}^{M} \beta_m x_{i,\kappa}^m + \frac{1}{g_{i,\kappa}} \sum_{m=1}^{M} \sum_{j=1}^{n_\kappa} \gamma_m g_{ij,\kappa} x_{j,\kappa}^m + \eta_\kappa + \varepsilon_{i,\kappa}$$

$$\varepsilon_{i,\kappa} = \mu g_{i,\kappa} + \phi \sum_{j=1}^{n_\kappa} g_{ij,\kappa} \varepsilon_{j,\kappa} + \upsilon_{i,\kappa}$$

1 Standard OLS

- A substantial part of the variance that is not explained.
- Strong evidence of spatial correlation in the residuals.

2 ML

$\hat{\mu}$, $\hat{\phi}$	Positive and highly statistically significant
$b_i(g, \Phi)$	Range: [0.32 , 3.48]; Average: 1.65; Standard deviation: 2.79
$\hat{\mu}/\hat{\phi}$	Statistically significant and non-negligible in magnitude • +1 standard deviation K-B → +7% Education outcome • +1 Parental education → +17%

$$z_i^* (\mathbf{g}) = \frac{\mu}{\phi} b_i (\mathbf{g}, \phi)$$

Part 4

Extension Ⅰ
Alternative Formulations

4.1 Alternative Formulations

[original] two efforts y_i , z_i

3 models: one effort → equilibrium effort = weighted Katz-Bonacich centrality index

achievement: network location & individual idiosyncratic characteristics ?

[original] both efforts contribute positively

if negatively: $y_i^*(\mathbf{x}, \mathbf{g}) = y_i^{0*}(\mathbf{x}) - z_i^*(\mathbf{g})$ (7') $\quad y_i^*(\mathbf{x}, \mathbf{g}) = \theta_i(\mathbf{x}) - \frac{\mu}{\phi} b_i(\mathbf{g}, \phi)$ (8')

sign of μ (empirical + ; 6% −)

4.2 Alternative Measures of Unit Centrality

1. Degree centrality

$$\delta_i(\mathbf{g}_\kappa) = g_i = \sum_{j=1}^{n} g_{ij} \qquad \delta_i^*(\mathbf{g}_\kappa) = \frac{g_i}{n_\kappa - 1} = \frac{\sum_{j=1}^{n} g_{ij}}{n_\kappa - 1}$$

m - 0.35
sd - 0.18

2. Closeness centrality

$$c_i(\mathbf{g}_\kappa) = \frac{1}{\sum_j d_{ij,\kappa}} \qquad c_i^*(\mathbf{g}_\kappa) = \frac{n_\kappa - 1}{\sum_j d_{ij,\kappa}}$$

m - 0.49
sd - 0.27

3. Betweenness centrality

$$f_i(g_\kappa) = \sum_{j<l} \frac{\text{numberofshortestpathsbetween} j \text{and} l \text{through} i \text{in} \mathbf{g}_\kappa}{\text{numberofshortestpathsbetween} j \text{and} l \text{in} \mathbf{g}_\kappa}$$

$$f_i^*(g_\kappa) = \frac{f_i(g_\kappa)}{(n_\kappa - 1)(n_\kappa - 2)/2}$$

m - 0.45
sd - 0.51

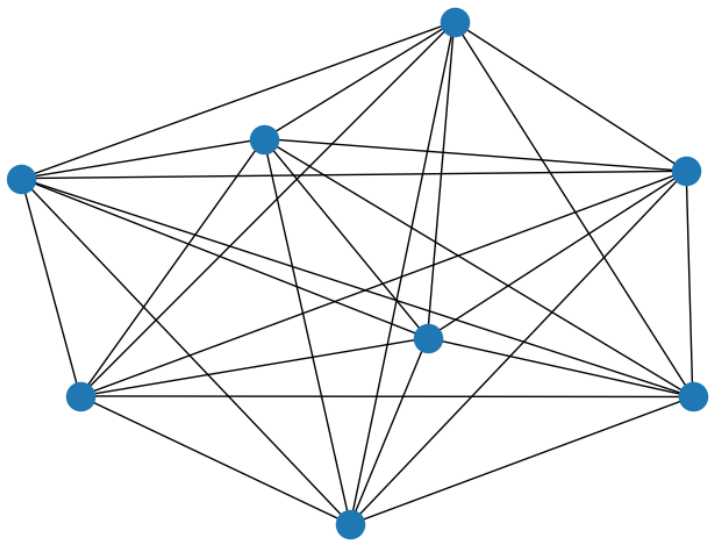
Part 5

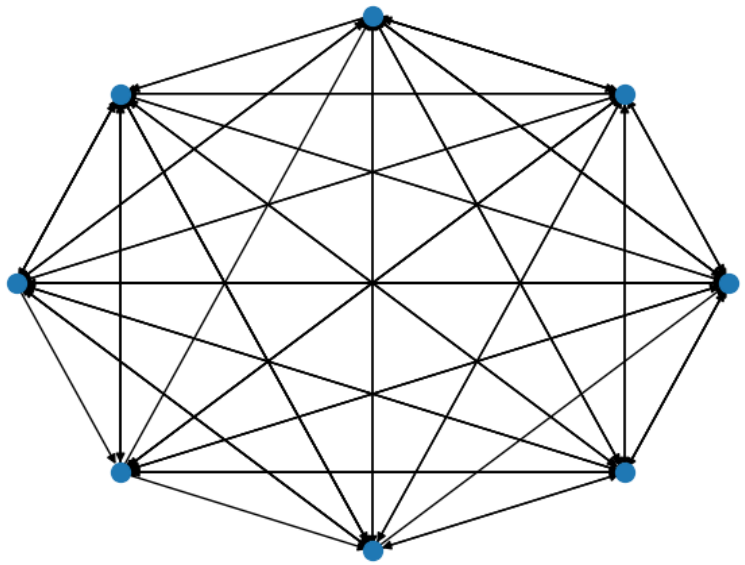
Extension Ⅱ
Directed Networks

5.1 Directed Networks

Assumption:

- Relationships reciprocal ≥ directed
- $g_{ij,k} = g_{ji,k} \geq g_{ij,k} \neq g_{ji,k}$
- 14% relationships are not reciprocal

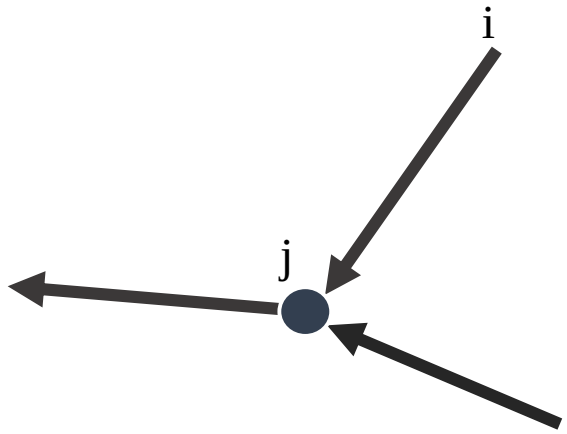




5.2 Difference

Theoretically :

- $g_{ij,k} \neq g_{ji,k}$ G： Asymmetric
- $\omega(g)$： Spectral radius of the adjacency matrix G
- $\phi\omega(g) < 1$， Proposition 1 holds true



Empirically：

- still statistically significant
- Slightly lower in magnitude(5.6 vs 7%)

TABLE 3

Model (9): Maximum Likelihood estimation results
Dependent variable: school performance index

	Undirected networks	Directed networks
Number of best friends (μ)	0·0314**	0·0323**
	(0·0149)	(0·0152)
Peer effects (ϕ)	0·5667***	0·5505***
	(0·1433)	(0·1247)
Individual socio-demographic variables	yes	yes
Family background variables	yes	yes
Protective factors	yes	yes
Residential neighbourhood variables	yes	yes
Contextual effects	yes	yes
School-fixed effects	yes	yes
R^2	0.8987	0.8905

6 总结

1 本文建立了一个关于同伴效应的网络博弈模型，求解其纳什均衡

$$y_i^*(\mathbf{x},\mathbf{g}) = \underbrace{y_i^{0*}(\mathbf{x})}_{\text{idiosyncratic}} + \underbrace{z_i^*(\mathbf{g})}_{\text{peer effect}} = \sum_{m=1}^{M}\beta_m x_i^m + \frac{1}{g_i}\sum_{m=1}^{M}\sum_{j=1}^{n}\gamma_m g_{ij}\, x_j^m + \frac{\mu}{\phi}\mathbf{b}(\mathbf{g},\phi)$$

从理论上说明了个体行为会受到其所处网络的影响，并明确最终均衡结果与博纳西克中心性有关

2 本文利用 Add Health 数据库关于学生网络关系、个人特性、学业成绩等信息的详细数据库，从实证角度检验了该模型。其中重要参数 μ 、 ϕ 都是显著的，说明该模型具有较好解释力

3 本文随后进行了两方面实证的拓展：

① 使用其他三种中心性指标：结果基本上都不显著

② 从无向网络扩展到有向网络：结果仍然是显著的

2020

THANKS

Appendix： A More General Model

$$u_i(\mathbf{y}^0, \mathbf{z}; \mathbf{g}) = \theta_i\, y_i^0 - \frac{1}{2}\left(y_i^0\right)^2 + \mu g_i z_i - \frac{1}{2} z_i^2 + \phi \sum_{j=1}^{n} g_{ij} z_i z_j$$

$$u_i(\mathbf{z}) = \alpha_i z_i + \frac{1}{2}\sigma_{ii} z_i^2 + \sum_{j \neq i} \sigma_{ij} z_i z_j.$$

$\alpha_i > 0$

$\sigma_{ii} < min\{0, \sigma_{ij}\ (j \neq i)\}$

$$assume\ \sigma_{ii} = \sigma < 0 \quad \because\ \tilde{u}_i(\mathbf{z}) = \frac{\sigma}{\sigma_{ii}} u_i(\mathbf{z}) = \frac{\alpha_i \sigma}{\sigma_{ii}} z_i + \frac{1}{2}\sigma z_i^2 + \sum_{j \neq i} \frac{\sigma_{ij}\sigma}{\sigma_{ii}} z_i z_j$$

$$= \tilde{\alpha}_i z_i + \tfrac{1}{2}\sigma z_i^2 + \textstyle\sum_{j \neq i} \tilde{\sigma}_{ij} z_i z_j$$

$$\boxed{\mathbf{z}^* = -\boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}}$$

$$\boldsymbol{F.O.C.} \quad \frac{\partial u_i}{\partial z_i} = \alpha_i + \sum_{j=1}^{n} \sigma_{ij}\, z_j = 0 \ \Rightarrow\ \boldsymbol{\alpha} + \boldsymbol{\Sigma}\mathbf{z} = \mathbf{0}$$

$$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)^T \quad \boldsymbol{\Sigma} = [\sigma_{ij}]$$

Additive Decomposition of the Interaction Matrix

$$\underline{\sigma} = \min\{\sigma_{ij} \mid i \neq j\} \qquad \overline{\sigma} = \max\{\sigma_{ij} \mid i \neq j\}$$

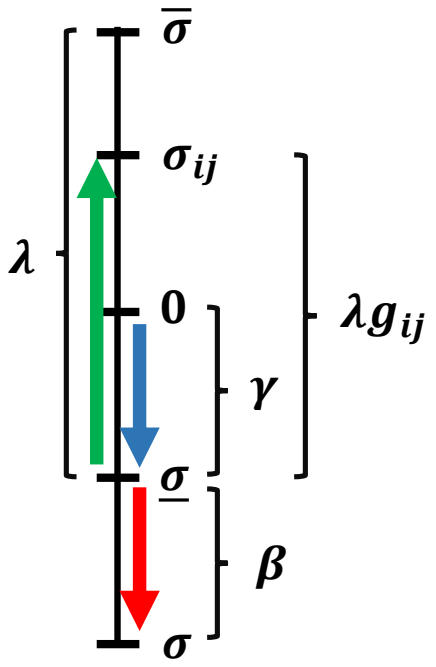
$$\gamma = -\min\{\underline{\sigma}, 0\} \geq 0 \qquad \lambda = \overline{\sigma} + \gamma \geq 0$$

$$g_{ij} = (\sigma_{ij} + \gamma)/\lambda \quad (i \neq j) \qquad \sigma = -\beta - \gamma \quad \beta > 0$$

$$g_{ii} = 0$$

$$\mathbf{\Sigma} = \boxed{-\beta \mathbf{I}}\boxed{-\gamma \mathbf{J}}\boxed{+\lambda \mathbf{G}}$$

$$\boldsymbol{I} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} \quad \boldsymbol{J} = \begin{bmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{bmatrix} \quad \mathbf{G} = [g_{ij}]$$



$$u_i(\mathbf{z}) = \alpha_i z_i + \frac{1}{2}\sigma_{ii} z_i^2 + \sum_{j \neq i} \sigma_{ij} z_i z_j.$$

$$\alpha_i = \mu\, g_i,\ \gamma = 0,\ \beta = 1,\ \lambda = \phi$$

$$\mu g_i z_i - \frac{1}{2} z_i^2 + \phi \sum_{j=1}^{n} g_{ij} z_i z_j$$

$$u_i(\mathbf{z}) = \alpha_i z_i - \frac{1}{2}(\beta - \gamma) z_i^2 - \gamma \sum_{j=1}^{n} z_i z_j + \lambda \sum_{j=1}^{n} g_{ij} z_i z_j, \quad \text{for all } i \in N.$$

Peer Effects and Kats-Bonacich Centrality

$$\mathbf{w_u}(\mathbf{g}, a) = \sum_{p=0}^{+\infty} a^p \mathbf{G}^p \mathbf{u} = (\mathbf{I} - a\mathbf{G})^{-1}\mathbf{u} \qquad \mathbf{b}(\mathbf{g}, a) = \mathbf{w}_{a\mathbf{G}\mathbf{1}}(\mathbf{g}, a)$$

$$-\boldsymbol{\Sigma}\mathbf{z}^* = \left[\beta\mathbf{I} + \gamma\mathbf{J} - \lambda\mathbf{G}\right]\mathbf{z}^* = \boldsymbol{\alpha}$$

$$\beta(I - \lambda/\beta\mathbf{G})\mathbf{z}^* + \gamma z^*\mathbf{1} = \boldsymbol{\alpha}$$

$$\beta\mathbf{z}^* = \left[\mathbf{I} - \lambda/\beta\mathbf{G}\right]^{-1}\left(\boldsymbol{\alpha} - \gamma z^*\mathbf{1}\right) = \mathbf{w}_{\boldsymbol{\alpha}}(\mathbf{g}, \lambda/\beta) - \gamma z^*\mathbf{w}_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)$$

$$\beta z^* = w_{\boldsymbol{\alpha}}(\mathbf{g}, \lambda/\beta) - \gamma z^* w_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)$$

$$z^* = \frac{w_{\boldsymbol{\alpha}}(\mathbf{g}, \lambda/\beta)}{\beta + \gamma w_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)}$$

$$\mathbf{z}^* = \frac{1}{\beta}\left[\mathbf{w}_{\boldsymbol{\alpha}}(\mathbf{g}, \lambda/\beta) - \frac{\gamma w_{\boldsymbol{\alpha}}(\mathbf{g}, \lambda/\beta)}{\beta + \gamma w_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)}\mathbf{w}_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)\right] \xrightarrow{\alpha_i = \alpha} \mathbf{z}^* = \frac{\alpha}{\beta + \gamma w_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)}\mathbf{w}_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)$$

$$\xrightarrow{\alpha_i = \mu g_i,\ \gamma = 0,\ \beta = 1,\ \lambda = \phi} \mathbf{z}^* = \frac{\mu}{\phi}\mathbf{b}(\mathbf{g}, \phi)$$

$$\boxed{z^* = \mathbf{1}^T\mathbf{z}^*}$$

$$\boxed{w_{\mathbf{1}}(\mathbf{g}, \lambda/\beta) = \mathbf{1}^T\mathbf{w}_{\mathbf{1}}(\mathbf{g}, \lambda/\beta)}$$

$$\boxed{w_{\boldsymbol{\alpha}}(\mathbf{g}, \lambda/\beta) = \mathbf{1}^T\mathbf{w}_{\boldsymbol{\alpha}}(\mathbf{g}, \lambda/\beta)}$$

Back