Spatial Interactions

J.S. Kim, E. Patacchini, P.M. Picard and Y. Zenou

Monash University
Cornell University
University of Luxembourg
Monash University

May 2019

Motivation

- Social interactions and networks are a critical engine for the economic growth of nations and regions (Romer (1986); Lucas (1988)).
- Glaeser (2000): the existence of cities critically hinges on how social interactions and networks can be facilitated across the space of urban entities.
- Yet, very few studies look at the interactions between the social space and the geographical space.

Motivation

- Here: Provides a bridge between two literatures: the traditional urban models and the recent social network models.
- While there is a common recognition of cities as a major places for social interactions, traditional urban models do not consider the presence of social interactions and social capital.
- Network papers (usually) assume that the existence and intensity of dyadic contacts do not depend on location, i.e. do not consider the geographical location of the agents.

Urban economics and agglomeration: Theory

- Important literature in urban economics looking at how interactions between agents create agglomeration and city centers (Fujita and Thisse (2013), Duranton and Puga (2015).
- How spatial externalities affect the location of firms and households, urban density patterns, and productivity (Beckmann (1976), Ogawa and Fujita (1980), Lucas and Rossi-Hansberg (2002), etc.).
- Here: we take the urban conguration as given and explain how the location of each agent in the city affects his/her social interactions with other agents in the city.

Social networks: Theory

- Usually assume away agents' geographical location.
- Small literature that looks at both (Johnson and Gilles (2000), Brueckner and Largey (2008), Helsley and Strange (2007), Zenou (2013), Mossay and Picard (2011); Mossay et al. (2013), Helsley and Zenou (2014), Sato and Zenou (2015)).
- The social network is usually not directly related to the geographical location of agents.

Urban and social networks: Empirics

- Very small literature!
- Difficult to find detailed data on social contacts as a function of geographical distance between agents together with information on relevant socio-economic characteristics.
- Some indirect tests: Bayer et al. (2008), Hellerstein et al. (2011, 2014), etc.
- More recently: Bailey et al. (2018), Büchel et al. (2019) show that distance is highly detrimental to interpersonal exchange.

What we do

- Develop a simple model of social-tie formation where individuals care about the geographical location of other individuals.
- Analyze the relationship between geographical distance and social interactions.

What we do

- Develop a simple model of social-tie formation where individuals care about the geographical location of other individuals.
- Analyze the relationship between geographical distance and social interactions.
- Prediction of the model: level of social interactions is inversely related to the geographical distance.
- Travel costs and spatial dispersion of agents are barriers to the development of social capital formation.

What we do

- Develop a simple model of social-tie formation where individuals care about the geographical location of other individuals.
- Analyze the relationship between geographical distance and social interactions.
- Prediction of the model: level of social interactions is inversely related to the geographical distance.
- Travel costs and spatial dispersion of agents are barriers to the development of social capital formation.
- Because of the externalities that agents exert on each other, the equilibrium levels of social interactions and social capital are lower than the efficient ones.

What we do

- Structurally estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.

What we do

- Structurally estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.
- Find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size.

What we do

- Structurally estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.
- Find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size.
- These inefficiencies are the largest when the network is composed of 28 students and the smallest for 68 students.

What we do

- Structurally estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.
- Find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size.
- These inefficiencies are the largest when the network is composed of 28 students and the smallest for 68 students.
- The network that maximizes the average welfare should have 55 students whereas the one that maximizes social interactions should be of size 44.

What we do

- Structurally estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.
- Find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size.
- These inefficiencies are the largest when the network is composed of 28 students and the smallest for 68 students.
- The network that maximizes the average welfare should have 55 students whereas the one that maximizes social interactions should be of size 44.
- Subsidies on social interactions are more effective than subsidies on transportation costs.

The Model

- Linear city on the line segment $x \in [-b, b]$
- $\lambda(x) : [-b, b] \to R^+$: number of agents located at x.
- Unit mass population: $\int_{-b}^{b} \lambda(y) \mathrm{d}y = 1$.

The Model

- Linear city on the line segment $x \in [-b, b]$
- $\lambda(x) : [-b, b] \to R^+$: number of agents located at x.
- Unit mass population: $\int_{-b}^{b} \lambda(y)\mathrm{d}y = 1$.
- Each agent *visits every other agent* and benefits from social interactions.

The Model

- Utility from social interactions:

$$S(x) = \int_{-b}^{b} v\left(n(x,y)\right) s(y)\lambda(y)\mathrm{d}y$$

- $n(x,y)$: *frequency* of interactions that agent at x initiates with an agent at y who offers an interaction value $s(y)$.
- Assume that

$$v\left(n\left(x,y\right)\right) = n\left(x,y\right) - \frac{1}{2}\left[n\left(x,y\right)\right]^2.$$

The Model

- Interaction value offered by an agent residing at y:

$$s(y) = 1 + \alpha \int_{-b}^{b} n(y,z)s(z)\lambda(z)\mathrm{d}z$$

- Similar to eigenvector centrality.
- $\alpha > 0$: importance of others' social capital in an agent's social capital formation.
- $s(y)$ as the social capital of the agent located at y.

The Model

- Each agent located at x incurs a cost of visiting another agent residing at y equal to: $c(x-y)$.
- Cost is symmetric and increases with distance $|x-y|$: $c(z) = c(-z)$ and $c'(z) > 0 \; \forall z > 0$.
- Total social interaction cost of an agent located at x:

$$C(x) = \int_{-b}^{b} n(x,y)c(x-y)\lambda(y)\mathrm{d}y$$

Social capital and space

- Utility:

$$U(x) = S(x) - C(x)$$

$$= \int_{-b}^{b} \{v\,(n(x,y))\,s(y) - n(x,y)c(x-y)\}\,\lambda(y)\mathrm{d}y$$

- Define the *access cost measure* as:

$$g(y) \equiv \int_{-b}^{b} c(y-z)\lambda(z)\mathrm{d}z$$

and

$$s_0 \equiv \frac{1-\alpha^2 \int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z}{1-\alpha},$$

Proposition

Assume $0 < \alpha < 1$ *and* $s_0 - \alpha \left[\max_y g(y)\right] > c(2b)$. *Then, there exists a unique equilibrium* $(n^*(x,y), s^*(y))$, *defined for all* x, y, *such that*

$$n^*(x,y) = 1 - \frac{c(x-y)}{s^*(y)}$$

and

$$s^*(y) = s_0 - \alpha g(y)$$

- Condition assumes away corner solutions and assume *global interactions* so that agents interact with every other agent in the city.
- For individual x, the number of interactions $n^*(x, y)$ between x and y increases with y's social capital
- $n^*(x, y)$ decreases with the distance between x and y.

Proposition

Lower travel costs increase social capital for all agents. An increase in α, the importance of peers' social links, increases each agent's social capital for small enough travel cost.

- A rise in α has ambiguous effects.
- An agent's social capital increases with higher α because she places greater value on the social capital of her interaction partners
- and because her partners themselves have higher social capital.
- However, as α increases, she reduces her frequency of interactions with the partners with higher social capital, which reflects a *substitution effect* between the *frequency* and the *quality* of social interactions.

Proposition

Suppose linear or convex travel cost functions. Then,

- (*i*) *A mean preserving increase in the spread of a symmetric distribution* λ *decreases social capital for all agents;*
- (*ii*) *Social capital is less spatially dispersed than agents if* $x^2\lambda(x) / \int z^2\lambda(z)\mathrm{d}z$ *is a mean preserving spread of the distribution of* $\lambda(x)$ *around its mean* $x = 0$.

- Provided that travel costs have appropriate regularity properties, a larger spatial dispersion of agents reduces the social capital in the city and social capital is less spatially dispersed than the agents.

Uniform distribution of agents and linear travel costs

- Assume uniform distribution of agents in the city: $\lambda(x) = 1/2b$. Thus $\int_{-b}^{b} \lambda(y)\mathrm{d}y = 1$.
- Assume linear travel costs: $c(x) = c_1 |x|$ where $c_1 > 0$. Then,

$$n^*(x,y) = 1 - \frac{c\,|x-y|}{s^*(y)}$$

- and

$$s^*(y) = 1 + \frac{\alpha}{2b} \int_{-b}^{b} n^*(y,z)s^*(z)\mathrm{d}z$$

Efficient social interactions

- The planner chooses the profiles of social interactions $n(\cdot,\cdot)$ and social capital $s(\cdot)$ that maximize the aggregate utility

$$W = \int_{-b}^{b} U(x)\lambda(x)\mathrm{d}x = \int_{-b}^{b} \left[S(x) - C(x)\right]\lambda(x)\mathrm{d}x$$

subject to the social capital constraint

$$s(x) \leq 1 + \alpha \int_{-b}^{b} n(x,z)s(z)\lambda(z)\mathrm{d}z$$

Efficient social interactions

- Equilibrium is inefficient.
- Main externality for social interactions:
- When the planner chooses $n(x, y)$, she considers both the benefit and cost to agent x but also the fact that an increase in x's social capital increases y's social capital.
- This latter effect is not considered by agent x in equilibrium.

Efficient social interactions

- Equilibrium is inefficient.
- Main externality for social interactions:
- When the planner chooses $n(x, y)$, she considers both the benefit and cost to agent x but also the fact that an increase in x's social capital increases y's social capital.
- This latter effect is not considered by agent x in equilibrium.
- The weight that the planner puts on raising another agent's social capital increases with the importance of interactions, α, and with the social benefit of relaxing the social capital constraint, $\chi(x)$.
- $\chi(x)$ is the Kuhn-Tucker multiplier of the social capital constraint. It measures the welfare value of a marginal increase of the social capital of an agent located at x.

Efficient social interactions

Proposition

The equilibrium frequency of interactions and level of social capital are lower than the efficient ones.

- The planner internalizes the effect that each agent has on others' social capital when she entertains more intense social interactions.
- Thus, the planner imposes agents to increase their frequency of social interactions above the equilibrium level.

Efficient social interactions

- Can the efficient allocation of social interactions be decentralized with subsidies $\sigma(x,y)$ and $\tau(x,y)$ for social interactions and travel costs?
- The utility becomes:

$$
\begin{aligned}
U(x) &= S(x) - C(x) \\
&= \int_{-b}^{b} \{ v(n(x,y)) \left[s(y) + \sigma(x,y) \right] \\
&\quad - n(x,y) \left[c(x-y) - \tau(x,y) \right] \} \lambda(y) \mathrm{d}y
\end{aligned}
$$

Efficient social interactions and subsidies

Proposition

- *The first best solutions* $n^o(x,y)$ *and* $s^o(x)$ *can be restored by setting* $\sigma(x,y) = 0$*, i.e. social interactions should not be subsidized, and* $\tau(x,y) = \alpha\chi^o(x)s^o(y)$*, i.e. trips should be subsidized as a function of the locations of the destination and origin partners.*
- *The subsidy* $\tau(x,y)$ *should be higher for trips to partners who have higher social capital and for trips from partners whose social capital increases more with additional interactions.*

Efficient social interactions and subsidies

- The planner does not subsidize the agents with high social capital but only subsidizes the trips to these agents.
- This result contrasts with Helsley and Zenou (2014) (with two location points), who advocate that the planner should subsidize the most central agents.
- Here, decentralization would be difficult to implement because subsidies depend on both the origins and destinations of social interactions

Empirical strategy

- To bring the model to the data, we need to introduce agents' heterogeneity.
- We assume that the benefits of the intensity of interactions between individuals x and y also depends on their social distance, that is on their distance in terms of socio-demographic characteristics:

$$v\left(n(x,y)\right) = \left(n_0 + \theta(x,y)\right) n\left(x,y\right) - \frac{1}{2}\left[n\left(x,y\right)\right]^2,$$

- where $\theta(x,y)$ denotes the social distance between x and y and n_0 is a positive constant.

Empirical strategy

- Data from R networks ($r = 1, ..., R$), each comprised of N_r agents.
- Individual i resides in location x, individual j in location y, and individual k in location z.
- The geographic distance between individuals i and j is denoted by $d_{ij,r}$.
- $\theta_{ij,r}$ denotes the social distance (gender, race) between i and j in network r.
- Equilibrium values (n_0 positive constant):

$$n^*_{ij,r} = n_0 - \frac{c d_{ij,r}}{s^*_{j,r}} + \theta_{ij,r}$$

$$s^*_{j,r} = 1 + \frac{\alpha}{2b_r} \sum_{k=1}^{N_r} n^*_{jk,r} s^*_{k,r}$$

Empirical strategy

- Social distance to depend on observed (pair-level) individual characteristics $x_{ij,r}$ and on unobserved factors $\varepsilon_{ij,r}$.
- Undirected network:

$$\theta_{ij,r} = \sum_{m=1}^{M} \beta_m |x_{i,m,r} - x_{j,m,r}| + \sum_{m=1}^{M} \beta_{M+m}(x_{i,m,r} + x_{j,m,r}) + \varepsilon_{ij,r}$$

- Directed network:

$$\theta_{ij,r} = \sum_{m=1}^{M} \beta_m (x_{i,m,r} - x_{j,m,r}) + \sum_{m=1}^{M} \beta_{M+m}(x_{i,m,r} + x_{j,m,r}) + \varepsilon_{ij,r}$$

Estimation

- Besides agents' characteristics $x_{ij,r}$, the data provide:
- $n^*_{ij,r}$, the intensity of social interactions between agents i and j in network r
- $d_{ij,r}$, the geographical distance between agents i and j in network r.
- $2b_r$, the maximum geographical distance between two agents in network r, i.e. $2b_r = \max d_{ij,r}$.
- Using this information, we need to recover α, β, c, n_0, and the equilibrium social capital, $s^*_{j,r}$.

Estimation

- Besides agents' characteristics $x_{ij,r}$, the data provide:
- $n^*_{ij,r}$, the intensity of social interactions between agents i and j in network r
- $d_{ij,r}$, the geographical distance between agents i and j in network r.
- $2b_r$, the maximum geographical distance between two agents in network r, i.e. $2b_r = \max d_{ij,r}$.
- Using this information, we need to recover α, β, c, n_0, and the equilibrium social capital, $s^*_{j,r}$.
- Method of simulated moments (MSM) proposed by McFadden (1989) and Pakes and Polard (1989).
- The objective of MSM estimation is to find the parameter vector that provides the simulated level of social interactions that best matches the observed level of social interactions.

Estimation procedure

- We define

$$\mathbf{s}^*(\mathbf{X}_r, \boldsymbol{\Delta}_r, b_r, \mathcal{E}_r; \mu)$$

$$\equiv \left[\mathbf{I}_{N_r} - \tfrac{\alpha}{2b_r}\left(\mathbf{N}_0 + \boldsymbol{\Theta}_r\right)\right]^{-1} \left(\mathbf{I}_{N_r} - \tfrac{\alpha c}{2b_r}\boldsymbol{\Delta}_r\right) \mathbf{1}_{N_r}$$

- $\boldsymbol{\Delta}_r = \left(d_{ij,r}\right)$, $\boldsymbol{\Theta}_r = \left(\theta_{ij,r}\right) = (x_{ij,r}^T \beta + \varepsilon_{ij,r})$.
- $\mathbf{N}_0$ is an $N_r \times N_r$ matrix in which the off-diagonal elements are n_0, and the diagonal elements are zero.
- $\mu = (n_0, \alpha, c, \beta^{\mathrm{T}}, \sigma_\varepsilon^2)^{\mathrm{T}}$ denotes the vector of all parameters where σ_ε^2 is the variance of ε.
- Let $s_j^*(\mathbf{X}_r, \boldsymbol{\Delta}_r, b_r, \mathcal{E}_r; \mu)$ be the jth element of $\mathbf{s}^*(\mathbf{X}_r, \boldsymbol{\Delta}_r, b_r, \mathcal{E}_r; \mu)$,

Estimation procedure

- By using $s_j^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \mu)$, we obtain

$$n_{ij,r}^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \mu) = n_0 - \frac{cd_{ij,r}}{s_j^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \gamma)} + x_{ij}^{\mathrm{T}}\beta + \varepsilon_{ij}.$$

Estimation procedure

- We draw T sets of simulation errors $\varepsilon_{ij,r}^{(t)}$, $t = 1, \cdots, T$ for all pairs and all networks.
- Next, we compute social capital $\mathbf{s}^{(t)}$ and predict the intensity of social interactions $\hat{n}_{ij,r}^{(t)}$ for each set of errors using the equations defining $\mathbf{s}^*(\mathbf{X}_r, \boldsymbol{\Delta}_r, b_r, \mathcal{E}_r; \mu)$ and $n_{ij,r}^*(\mathbf{X}_r, \boldsymbol{\Delta}_r, b_r, \mathcal{E}_r; \mu)$.
- Then, the prediction error is given by

$$\begin{aligned}\hat{v}_{ij,r} &= n_{ij,r}^* - \frac{1}{T}\sum_{t=1}^{T}\hat{n}_{ij,r}^{(t)} \\ &= n_{ij,r}^* - \frac{1}{T}\sum_{t=1}^{T}\left(n_0 - \frac{cd_{ij,r}}{\hat{s}_j(\mathbf{X}_r, \boldsymbol{\Delta}_r, b_r, \mathcal{E}_r^{(t)}; \mu)} + x_{ij}^{\mathrm{T}}\beta + \varepsilon_{ij,r}^{(t)}\right),\end{aligned}$$

where $\mathcal{E}_r^{(t)}$ is the matrix of the tth set of simulation errors.

Estimation procedure

- The prediction error $\hat{v}_{ij,r}$ is uncorrelated with exogenous data $x_{ij,r}$ and $d_{ij,r}$ at the true parameter value μ_0.
- That is,

$$E(\hat{v}_{ij,r}|\mathbf{X}_r, \boldsymbol{\Delta}_r; \mu = \mu_0) = 0$$

- We have

$$E(\hat{v}_{ij,r}) = 0, E(\hat{v}_{ij,r}x_{ij,r}) = 0 \text{ and } E(\hat{v}_{ij,r}d_{ij,r}) = 0.$$

Estimation procedure

- From this, we can construct $(2M + 2)$ moment conditions:
- $E(\hat{v}_{ij,r} x_{ij,r}) = 0$ contains $2M$ moment conditions since x_{ij} is a $2M \times 1$ vector of $|x_i - x_j|$'s and $(x_i + x_j)$'s.
- $E(\hat{v}_{ij,r}) = 0$ is one moment condition.
- $E(\hat{v}_{ij,r} d_{ij,r}) = 0$ is one moment condition.

Estimation procedure

- We have constructed $(2M+2)$ moment conditions but we have $(2M+4)$ parameters to estimate:
- M is the number of the x variables, and we have $|x_i - x_j|$ and $(x_i + x_j)$, so we have $2M$ coefficients on those pair-level variables.
- $+3$: three structural parameters: α, c, n_0.
- $+1$: the variance of unobserved variable: σ_ϵ^2.
- So the model is still under-identified.

Estimation procedure

- To ensure identification, we utilize the relation between social capital and the eigenvector centrality of social interactions matrix $\mathbf{N}_r^*$.
- Social capital equation:

$$s_{j,r}^* = 1 + \frac{\alpha}{2b_r} \sum_{k=1}^{N_r} n_{jk,r}^* s_{k,r}^*, \tag{21}$$

- Eigenvector centrality $EC_{j,r}$:

$$EC_{j,r} = \frac{1}{\lambda} n_{jk,r}^* EC_{k,r},$$

- where λ is the largest eigenvalue of $\mathbf{N}_r^*$.

Estimation procedure

- The eigenvector centrality from $\mathbf{N}_r^*$ and the eigenvector centrality from the predicted social interaction matrix, $\hat{\mathbf{N}}_r^*$, must be close to each other.
- We define another $N_r \times 1$ vector of predicted errors $\hat{\xi}_r$ such that

$$\hat{\xi}_r = EC_r - \frac{1}{T}\sum_t \widehat{EC}_r^{(t)},$$

- where $\widehat{EC}_r^{(t)}$ is the eigenvector centrality corresponding to the predicted social interactions network $\hat{\mathbf{N}}_r^{*,(t)}$ with respect to the tth simulation errors.

Estimation procedure

- These prediction errors are mean independent of $x_{ij,r}$ and $d_{ij,r}$ at the true parameter value μ_0.
- That is,
$$E(\hat{\xi}_r|\mathbf{X}_r, \mathbf{\Delta}_r; \mu = \mu_0) = 0$$
- We have:
$$E(\hat{\xi}_r) = 0, E(\hat{\xi}_r x_{ij,r}) = 0 \text{ and } E(\hat{\xi}_r d_{ij,r}) = 0$$
- We have $(2M + 2)$ additional moment conditions.
- Thus a total of $(4M + 4)$ moment conditions for $(2M + 4)$ parameters, and the model is identified (over-identified).

Data

- Database on friendship networks from the National Longitudinal Survey of Adolescent Health (Add Health).
- School-based survey which contains extensive information on a representative sample of students who were in in grades 7-12 in 1995.
- Three key features:
  (i) the nomination-based friendship information,
  (ii) the detailed information about the intensity of social interactions between each of two friends in the network;
  (iii) the geo-coded information on residential locations, which allows us to measure the geographical distance between individuals.

Friendship nomination

- The friendship information is based upon actual friend nominations at school.
- All students who were present at school in the interview day received the questionnaire.
- Pupils were asked to identify their best school friends from a school roster (up to five males and five females).
- The limit in the number of nominations is not binding (even by gender).

Frequency and nature of interactions

- Frequency and nature of interaction with each nominated friend j.
- "Did you go to {NAME}'s house during the past seven days?";
- "Did you meet {NAME} after school to hang out or go somewhere during the past seven days?";
- "Did you spend time with {NAME} during the past weekend?";
- "Did you talk to {NAME} about a problem during the past seven days?";
- "Did you talk to {NAME} on the telephone during the past seven days?".
- n_{ij} is measured by summing all these items (only 0,1 answers) so that the maximum value of n_{ij} is 5 and the minimum is 0.

Residential locations

- The geographical location of each house is recorded.
- Latitude and longitude coordinates are calculated for each home address and then translated into $X-$ and $Y-$coordinates in an artificial space.
- We use this information to derive the spatial distance between students.
- The maximum geographical distance between two students, which is calculated for each network separately, is about 173 kilometers.
- The average distance is about 8.8 kilometers, while the median distance is about 5.3 kilometers.

Data

- Our final sample consists of about 900 individuals distributed over roughly 100 networks.
- Network size: from 4 to 70 members.
- 58% are female and 20% are blacks. Slightly more than 70% live in a household with two married parents.
- The average parental education is high school graduate.
- The performance at school, as measured by the grade point average or GPA, exhibits a mean of 2.98, meaning slightly less than a grade of “B”.
- The average family income is 44,562 in 1994 dollars.

Empirical results

- Structural parameters (all statistically significant).
- The baseline level of social interactions n_0 is roughly $2.2 - 3.4$.
- The estimated cost of transportation is $0.00011 - 0.00026$ across specifications.
- Combined with average pairwise distance (8.78 kilometers), the average estimated transportation cost is $0.001 - 0.0023$.
- Finally, α (importance of others' social capital in an agent's social capital formation) is equal to $0.0167 - 0.0334$.

Empirical results

- Also, strong homophily behaviors in terms of race and gender (estimation of the βs).
- A pair of females is associated with 0.0096 more social interactions than a pair of males.
- Students are preferred as social interaction partners if they are female, black, older students (higher grade), are physically more developed, are more religious and have better educated parents.

Dispersion

- Test Proposition (without using the structural estimates): a mean preserving increase in the spread of a symmetric distribution λ decreases social capital for all agents.
- Is there is a *negative* relationship between $\overline{n}_r^*$ (average interaction in each network) and $\overline{d}_r$ (average distance between student pairs in each network)?
- Is there is a *negative* relationship between $\overline{s}_r^*$ (average social capital in each network) and $\overline{d}_r$?
- Measure dispersion (DISP) as the average distance between individuals and their network baricenter, $\overline{d}_r = \frac{1}{N_r} \sum_i d_{i,r}$.

Dispersion

- We test:

$$\overline{n}_r^* = \gamma_0 + \gamma_1 N_r + \gamma_2 \left(N_r\right)^2 + \gamma_3 \overline{d}_r + \gamma_z z_r + \gamma_x x_r + \epsilon_r$$

$$\overline{s}_r^* = \delta_0 + \delta_1 N_r + \delta_2 \left(N_r\right)^2 + \delta_3 \overline{d}_r + + \gamma_z z_r + \delta_x x_r + \zeta_r$$

- where N_r is the size (in terms of population) of network r, z_r are network measures and x_r are network-level socio-economic control variables (such as average family income in network r, etc.).

Table: Social interactions and geographic dispersion of students

	Average social interactions				Average social capital			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Disp	-0.0913***		-0.120***		-0.0052***		-0.0055***	
	(0.0327)		(0.0291)		(0.0011)		(0.0012)	
Av. dis		-0.0661***		-0.0819***		-0.0036***		-0.0037***
		(0.0231)		(0.0204)		(0.0008)		(0.0008)
Pop			0.146***	0.143***			0.0013	0.0011
			(0.0315)	(0.0357)			(0.0008)	(0.0008)
Pop^2			-0.0016***	-0.0016**			-0.000006	-0.000004
			(0.00047)	(0.00063)			(0.00001)	(0.00001)
Obs	104	104	104	104	104	104	104	104
R^2	0.064	0.074	0.308	0.312	0.245	0.259	0.306	0.314

Note: Dispersion of a network is measured by taking the average of distances from each student's home to the network center. Average distance is the average of pairwise distances of students in a network.

Dispersion

- A one kilometer increase in the geographical dispersion of individuals is associated with an approximately 0.06–0.07 decrease (5–6% decrease relative to the mean) in the average social interactions
- and a 0.002 decrease (0.2% decrease relative to the mean) in the average social capital.

Dispersion

- Non-monotonic relationship between the average interactions in a given network $\overline{n}_r^*$ and the network size N_r.
- Using Column (4), we have:

$$\frac{\partial \overline{n}_r^*}{\partial N_r} = \gamma_1 + 2\gamma_2 N_r = 0.143 - 2(0.0016)N_r = 0$$

- Solving this equation leads to: $N_r = \frac{0.143}{2(0.0016)} \approx 44$.
- $N_r = 44$ is thus the size of the network that *maximizes* average social interactions in our data.

Welfare

- From structural estimations: estimated values of n_0, α, c and $\theta_{ij,r}$.
- From the data, we know b_r and $d_{ij,r}$.
- By plugging these values into the first-order conditions of the planner program , we can solve *numerically* these equations and determine $n^o_{ij,r}$, for each pair i, j, $s^o_{j,r}$ for all j, and $v_{i,r}$ for all i.
- According to the Proposition on welfare, we should find that students socially interact too little compared to the social optimal, i.e. $n^o_{ij,r} > n^*_{ij,r}$, $\forall i, j$, and $s^o_{j,r} > s^*_{j,r}$, $\forall i_{xr}$.

Welfare

Table: Social interactions: Optimal level vs. observed level

Social interactions				
Optimal level	Observed level	Average difference	Minimum difference	Maximum difference
2.349	1.144	1.205	-0.810	2.741

Note: The statistics are computed using the network-level average social interactions and social capital from 104 networks. For example, the largest difference between the average levels of optimal and observed social interactions is 2.741. Note that these statistics differ from pair-level averages.

Welfare

Table: Social capital: Optimal level vs. observed level

Social capital				
Optimal level	Observed level	Average difference	Minimum difference	Maximum difference
1.265	1.028	0.236	-0.095	3.504

Note: The statistics are computed using the network-level average social interactions and social capital from 104 networks. For example, the largest difference between the average levels of optimal and observed social capital is 3.504. Note that these statistics differ from pair-level averages.

Welfare

- On average, each pair interacts 1.2 fewer times than is socially optimal.
- The difference between the socially optimal and the observed levels of social interactions varies from -0.81 to 2.74 across networks.
- Although there are a few networks where the observed level is larger than the optimal level, most networks' interactions fall short of the optimum..
- Students also have less social capital than optimal (by 0.236, or approximately 25%, on average).

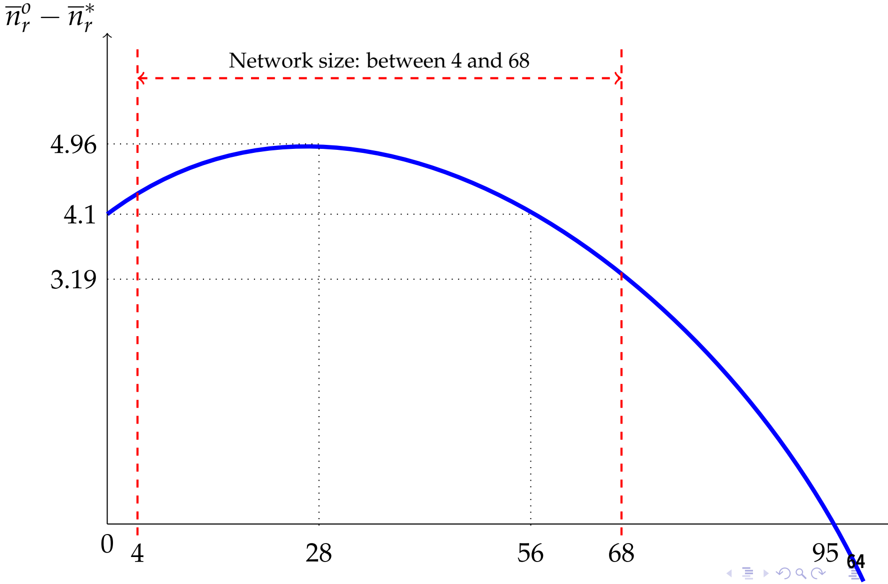
Network size and efficiency of social interactions

- We would like to find which variables are closely associated with the discrepancy between the optimal level and the observed level.
- We regress:

$$\overline{n}_r^o - \overline{n}_r^* = \gamma_0 + \gamma_1 N_r + \gamma_2 \left(N_r\right)^2 + \gamma_3 \overline{d}_r + \gamma_z z_r + \gamma_x x_r + \epsilon_r.$$

$$\overline{s}_r^o - \overline{s}_r^* = \delta_0 + \delta_1 N_r + \delta_2 \left(N_r\right)^2 + \delta_3 \overline{d}_r + \delta_z z_r + \delta_x x_r + \zeta_r.$$

Network size and inefficiency



Network size and efficiency of social interactions

- We find a non-monotonic relationship between $\overline{n}_r^o - \overline{n}_r^*$ and N_r (hump-shaped relationship).
- $\overline{n}_r^o - \overline{n}_r^*$ and N_r increases until the network size reaches (approximately) 28 students and then decreases.
- $N_r = 28$ is the size of the network that *maximizes* these inefficiencies.
- $N_r^{max} = 68$ is the size of the network that *minimizes* these inefficiencies.

Network size and average welfare

- Another interesting exercise, for which do not have theory, is to determine the optimal network, (i.e. the one that maximizes total welfare).
- We regress:

$$AW_r^* = \delta_0 + \delta_1 N_r + \delta_2 (N_r)^2 + \delta_z z_r + \delta_x x_r + \epsilon_r$$

- where AW_r^* is average welfare per network.

Table: Optimal network design: average welfare and number of students

	(1) Welfare	(2) Welfare	(3) Welfare	(4) Welfare	(5) Welfare
Network population	-2.108*	-1.074	0.488	3.590	2.071
	(1.245)	(0.918)	(1.514)	(3.762)	(3.264)
Network population^2	0.0279	0.0114	-0.0040	-0.0402	-0.0187
	(0.0178)	(0.0133)	(0.0150)	(0.0394)	(0.0327)
Avg. geographic distance		-3.610**	-3.914*	-3.968*	-3.181*
		(1.755)	(2.117)	(2.148)	(1.834)
Avg. degree centrality			1.835	4.519	3.679
			(3.299)	(4.774)	(9.706)
Std.dev. of degree centrality			-8.068	-9.581	-4.249
			(9.719)	(9.643)	(10.05)
Avg. eigenvector centrality				87.22	158.5
				(116.4)	(194.9)
Clustering coefficient				3.164	76.42
				(97.79)	(198.0)
Diameter				-1.734	-0.805
				(1.521)	(2.489)

Network size and average welfare

- Average pairwise geographic distance is an important factor for designing an optimal network.
- The higher the distance between two individuals in a network, the lower the average welfare.

Policies

- Consider social-interaction subsidies and travel-cost subsidies that only target each individual but not a pair of individuals.

Subsidizing social interactions

- The planner subsidizes the intensity of social interactions $n(x,y)$ in the following way:

$$
\begin{aligned}
U(x) &= S(x) - C(x) \\
&= \int_{-b}^{b} \{v\,(n(x,y))\,s(y) - n(x,y)(x-y)\}\,\lambda(y)\mathrm{d}y \\
&+ \sigma \int_{-b}^{b} n(x,y)\lambda(y)\mathrm{d}y
\end{aligned}
$$

- where σ is the value of the social-interaction subsidy.
- Each individual x receives a fixed amount of money $\sigma \int_{-b}^{b} n(x,y)\lambda(y)\mathrm{d}y$ proportional to the individual x's social interaction effort with all her friends.

Subsidizing social interactions

- New equilibrium values $n^{\sigma}(x, y)$ and $s^{\sigma}(y)$.
- New total welfare per network.
- We find the subsidy σ_r^* that gives network r the same aggregate utility W_r^{σ} as the first best W_r^{o}.
- From the estimated value of the equilibrium model, we have α, c and n_0.
- From the data we have $d_{ij,r}$ and b_r.
- We can then numerically solve the new equilibrium equations defining $n^{\sigma}(x, y)$ and $s^{\sigma}(y)$.
- and the optimal subsidy that maximizes total welfare to obtain σ_r^*, $n_{ij,r}^{\sigma}$ and $s_{j,r}^{\sigma}$.

Subsidizing social interactions

Table: Policy levels for optimal outcomes

Subsidizing social interactions: σ		
Average	Minimum	Maximum
0.4133	-0.5473	10.3352

Note: The subsidy level for each network is computed for students in each network to obtain the optimal level of social interactions and social capital.

Subsidizing social interactions

- On average, a subsidy level of 0.4133 for each social interaction is required for a network to achieve the first best aggregate level of social interactions and social capital.
- Most networks are offered a positive subsidy, which reflects a lack of social interaction.
- We also compute a single subsidy σ^* for all networks, which allows individuals to achieve the first best as close as possible and we find $\sigma^* = 1.4534$.

Subsidizing transportation costs

- Marginal transport cost per distance is now equal to $c - \tau$ (τ subsidy on transportation costs).
- Total social interaction cost of an agent located at x is now given by

$$C(x) = \frac{1}{2b} \int_{-b}^{b} n(x,y)(c-\tau)\,|x-y|\,\mathrm{d}y$$

Subsidizing transportation costs

- As for the social interaction subsidy, we find the subsidy τ_r^* that gives the same aggregate utility W_r^τ in network r as the first best W_r^0.
- From the estimated value of the equilibrium model, we have α, c and $n_{0,r}$, and from the data $d_{ij,r}$ and b_r.
- We can then numerically solve the equlibrium equations and the optimal subsidy that maximizes Welfare to obtain τ_r^*, $n_{ij,r}^\tau$ and $s_{j,r}^\tau$.

Subsidizing transportation costs

Table: Policy levels for optimal outcomes

Subsidizing transportation costs: τ		
Average	Minimum	Maximum
0.9471	-0.2470	17.7655

Note: The subsidy level for each network is computed for students in each network to obtain the optimal level of social interactions and social capital.

Subsidizing transportation costs

- On average, a subsidy level of 0.9471 per kilometer is required for a network to achieve the first best aggregate level of social interactions and social capital.
- As above, most networks receive positive subsidies to entice more interactions.
- We also compute a single subsidy τ^* for all networks, which allows individuals to achieve the first best as close as possible, and we find $\tau^* = 5.8601$.

Comparing the two policies

- Compare these two policies at a given cost. Given that the planner has an amount T to spend, which policy should she choose?

Comparing the two policies

- Compare these two policies at a given cost. Given that the planner has an amount T to spend, which policy should she choose?
- Three different schemes.
- (i) Distribute the same amount $T_r = T/R$ for each network (uniform subsidy).
- (ii) Give an amount proportional to network population N_r so that $T_r = \frac{N_r}{\sum_{r'} N_{r'}} T$.
- (iii) Provide an amount proportional to the number of pairs $N_r(N_r - 1)$, i.e. $T_r = \frac{N_r(N_r-1)}{\sum_{r'} N_r'(N_{r'}-1)} T$.

Comparing the two policies

- Budget constraint for each policy.
- Social-interaction subsidy policy: Planner's budget constraint for each network r:

$$\frac{\sigma_r}{4b_r^2}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r} n_{ij,r} = T_r$$

- Transportation subsidy policy: Planner's budget constraint for each network r:

$$\frac{\tau_r}{4b_r^2}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r} n_{ij,r}d_{ij,r} = T_r$$

Comparing the two policies

Table: Comparison of two policies

Subsidy schemes	Networks that lead to higher welfare	
	Policy: σ	Policy: τ
(1) Uniform subsidy	81	11
(2) Propor to N_r	81	11
(3) Propor to $N_r(N_r - 1)$	97	7

(1) Uniform subsidy amount for each network, (2) Subsidy proportional to N_r, (3) Subsidy proportional to $N_r(N_r - 1)$.

Comparing the two policies

- Under the social-interaction subsidy policy, the total welfare is higher for most networks.
- If a planner has a given amount of money to spend, she should subsidize social interactions and not transportation costs because it yields greater improvements to total welfare.

Conclusion

- Analyze the relationship between geographical distance and social interactions.
- Prediction of the model: level of social interactions is inversely related to the geographical distance.
- Travel costs and spatial dispersion of agents are barriers to the development of social capital formation.
- Because of the externalities that agents exert on each other, the equilibrium levels of social interactions and social capital are lower than the efficient ones.

Conclusion

- Estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.

Conclusion

- Estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.
- Find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size.

Conclusion

- Estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.
- Find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size.
- These inefficiencies are the largest when the network is composed of 28 students and the smallest for 68 students.

Conclusion

- Estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.
- Find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size.
- These inefficiencies are the largest when the network is composed of 28 students and the smallest for 68 students.
- The network that maximizes the average welfare should have 55 students whereas the one that maximizes social interactions should be of size 44.

Conclusion

- Estimate the model using data on adolescents in the United States.
- Show that geographical distance is an hinder to social interactions.
- Find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size.
- These inefficiencies are the largest when the network is composed of 28 students and the smallest for 68 students.
- The network that maximizes the average welfare should have 55 students whereas the one that maximizes social interactions should be of size 44.
- Subsidies on social interactions are more effective than subsidies on transportation costs.

Policy implications

- Encouraging social interactions in cities are likely to enhance social welfare.

Policy implications

- Encouraging social interactions in cities are likely to enhance social welfare.
- Social mixing such as the Moving to Opportunity (MTO) programs in the United States.

Policy implications

- Encouraging social interactions in cities are likely to enhance social welfare.
- Social mixing such as the Moving to Opportunity (MTO) programs in the United States.
- Improve physical environment such as zoning laws and public housing rules.
- Glaeser (2000): individuals in larger apartment buildings are more likely to socialize with their neighbors.

Policy implications

- Encouraging social interactions in cities are likely to enhance social welfare.
- Social mixing such as the Moving to Opportunity (MTO) programs in the United States.
- Improve physical environment such as zoning laws and public housing rules.
- Glaeser (2000): individuals in larger apartment buildings are more likely to socialize with their neighbors.
- Stroebel et al. (2018): friendship networks are a mechanism that can propagate house price shocks through the economy via housing price expectations.

Policy implications

- Encouraging social interactions in cities are likely to enhance social welfare.
- Social mixing such as the Moving to Opportunity (MTO) programs in the United States.
- Improve physical environment such as zoning laws and public housing rules.
- Glaeser (2000): individuals in larger apartment buildings are more likely to socialize with their neighbors.
- Stroebel et al. (2018): friendship networks are a mechanism that can propagate house price shocks through the economy via housing price expectations.
- Important since social interactions can promote economic growth (for ex, information about jobs) because of the nonmarket intellectual spillovers that they generate