Spatial interactions[*]

Jun Sung Kim[†] Eleonora Patacchini[‡]

Pierre M. Picard[§] Yves Zenou[¶]

October 2021

Abstract

This paper studies how the formation of social ties are affected by the geographical location of other individuals and their social capital. We characterize the equilibrium in terms of both social interactions and social capital for a general distribution of individuals in the urban geographical space. We show that greater geographical dispersion decreases the average welfare from social interactions. We also show that the equilibrium frequency of interactions is lower than the efficient one. Using a unique geo-coded dataset of friendship networks among adolescents in the United States, we estimate the model and validate that agents interact socially less than the first best optimum. Our policy analysis suggests that, at the same cost, subsidizing social interactions yields a higher total welfare than subsidizing transportation costs.

JEL codes: D85, R1, Z13.

Key words: Social networks, location, structural estimation, policies.

[*] We are grateful to Jan K. Brueckner and Paco Maruhenda for insightful discussions.

[†] Kyung Hee University, South Korea. E-mail: junsungkim@khu.ac.kr.

[‡] Cornell University, USA, EIEF and CEPR. E-mail: ep454@cornell.edu.

[§] CREA, University of Luxembourg, Luxembourg, and CORE, Université catholique de Louvain, Belgium. Email: pierre.picard@uni.lu.

[¶] Monash University, Australia, and CEPR. E-mail: yves.zenou@monash.edu.

1 Introduction

Over the past two decades, the economics literature has increasingly utilized network analysis to understand decision-making.[1] Surprisingly, however, the importance of spatial proximity in the determination and intensity of network exchange remains under-examined. For Glaeser (2000), the existence of cities critically hinges on how social interactions and networks can be facilitated across the space of urban entities. However, traditional models in urban/spatial economics (Fujita, 1989) do not consider the presence of social interactions and social capital in cities. On the other hand, most papers from the network economics literature (Jackson, 2008) assume that the existence and intensity of dyadic contacts do not depend on agents' location.

In this paper, we develop a new theory of social-tie formation where individuals care about the geographical location of other individuals. In our model, a population of agents entertains social interactions in a unidimensional geographical space (the city). In this city, the fraction of individuals at each location is determined by a general distribution function. Each agent decides the frequency of her visits (social interactions) to every other agent in the city, and the value of each interaction depends on the social network of the visited agents. We define the value of such interactions as the *social capital* of the agent (Putnam, 2000). Social capital is thus defined in a recursive fashion: it increases with interactions with highly social individuals. When deciding how much to interact with others, agents face the following trade-off. Each agent can increase her social capital by interacting with highly social agents. However, social interactions requires costly travel to the other agents. We characterize the equilibrium in terms of social interactions and social capital for a general distribution of individuals in the geographical space. We show that a more spread spatial distribution of agents decreases the incentives to socially interact. We also show that the equilibrium frequencies of interactions are lower than the efficient ones. We demonstrate that a policy that subsidizes transportation costs can restore the first best but the subsidy should be higher for trips to individuals who have higher social capital and for trips from individuals whose social capital increases more with additional interactions.

We then estimate this model using data on patterns of social interactions among high school students in the US recorded in the National Longitudinal Survey of Adolescent Health

[1] For recent overviews, see Jackson (2008), Ioannides (2013), Jackson and Zenou (2015), Bramoullé, Rogers, and Galeotti (2016) and Jackson, Rogers, and Zenou (2017).

(Add Health). This survey contains information on friendship nominations, the strength of the interactions between friends, and also allows us to calculate the Euclidean distance between the homes of the respondents. Because residential decisions are taken by parents, this spatial distance is pre-determined to the friendship decisions of the children. Our main empirical challenges are due to the fact that the intensity of social interactions can be chosen at the same time as friends, and that the interaction value offered by a friend (social capital) is unobserved to the econometrician. We address these challenges by applying an indirect inference estimation method to simulate unobserved social capital. The main idea of this method is to simulate data from the model, which requires solving for the unobserved equilibrium social capital conditional on structural parameters and unobservables, in order to find the parameters for which the simulated data best match the observed data.[2]

The estimation results provide evidence supporting our theory. We find that transportation costs (and hence geographic distance), social distance, and combined levels of socio-demographic characteristics are all important factors in determining the intensity of social interactions. With the estimated model, we compute the planner's first best solution for the frequency of social interactions and compare it with the observed equilibrium level. Compared to the socially optimal level, our results show that students interact with each other far less and accumulate less social capital. We find that these inefficiencies can be explained by the size of the network. With the estimated model, we also simulate the level of social interactions after different policy interventions. By subsidizing social interactions or transportation costs, the policymaker can indeed improve the intensity of social interactions. At the same given cost, we find that subsidizing social interactions is more effective than subsidizing transportation costs because it leads to higher total welfare.

Our theoretical framework provides a bridge between two literatures: the traditional urban/spatial models and the recent social network models. There is an important literature in urban economics looking at how interactions between agents create agglomeration and city centers.[3] It is usually assumed that the level of the externality that is available to a particular agent depends on its location– the spillover is assumed to attenuate with distance – and on the spatial arrangement of economic activity. This literature (whose keystones include Beckmann, 1976; Ogawa and Fujita, 1980; Lucas and Rossi-Hansberg, 2002; Behrens,

[2] Fu and Gregory (2019) develop an equilibrium model of post-disaster neighborhood rebuilding choices with externalities and estimate the model using indirect inference to implement policy simulations.

[3] See Fujita and Thisse (2013) and Duranton and Puga (2015) for extensive literature reviews.

Duranton, and Robert-Nicoud, 2014; Helsley and Strange, 2014) examines how such spatial externalities influence the location of agents, urban density patterns, and productivity. For example, Glaeser (1999) develops a model in which random contacts influence skill acquisition, while Helsley and Strange (2014) consider a model in which randomly matched agents choose whether and how to exchange knowledge. Similarly, Berliant, Peng, and Wang (2002) show the emergence of a unique center in the case of production externalities. These models provide an interesting discussion of spatial issues in terms of use of residential space and formation of neighborhoods and show under which condition different types of city structures emerge. In this paper, we consider a different view. While the literature cited above aims at explaining different urban configurations (monocentric versus polycentric cities) and to derive conditions under which they emerge,[4] we take the urban configuration as given and explain how the location of each agent in the city affects her social interactions with other agents in the city. In other words, we simplify the urban configuration of the city but we open the black box of social interactions by examining how and why they form.

Regarding the vast literature on social networks, most of the papers looking at network formation assume away agents' geographical locations.[5] A small strand of the literature (Brueckner and Largey (2008), Helsley and Strange (2007), Zenou (2013), Mossay and Picard (2011, 2019), Helsley and Zenou (2014), Sato and Zenou (2015), Picard and Zenou (2018)) studies the role of social networks in cities but take the networks as given. In the current paper, link formation depends on the location of individuals in the geographical space. From an empirical point of view, studies in the economics literature on the relevance of geographical location for social interactions in networks are scarce (see Ioannides, 2013, for a survey). In fact, it is extremely 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 evidence can be found in Marmaros and Sacerdote (2006). Using data on email communication between Dartmouth college students, this paper shows that being in the same freshman dorm increases the volume of interactions by a factor of three.[6] Büchel

[4] For example, Ogawa and Fujita (1980), a prominent paper in this literature, consider a "locational potential function" in which a weighted average of pairwise Euclidean distances between firms has a negative effect on firms' profit. This acts as an agglomeration force for firms because it implies a (strictly) penalty cost for firm dispersion.

[5] Exceptions include Johnson and Gilles (2000) and Jackson and Rogers (2005). These studies, however, consider a framework where network formation is modeled on a link-by-link basis. As a result, it is impossible to fully characterize all the possible equilibria. See Jackson (2008) for a discussion of these issues.

[6] See also Fafchamps and Gubert (2007) who show that geographic proximity is a strong correlate of

and von Ehrlich (2019) measure social connectedness between postcode areas in Switzerland using mobile phone communication patterns between residents in different areas. They find that distance as measured by travel time is detrimental to private mobile phone interactions by exploiting an exogenous change in travel time.[7] Bailey et al. (2018b) and Bailey et al. (2020) reach a similar conclusion by using anonymized and aggregated data from Facebook to explore the spatial structure of social networks in the New York metropolitan area.

The vast literature in the computer science literature and statistical mechanics looking at the role of distance in social interaction uses primarily mobile phone data or online social networks data and is mainly concerned about describing the shape of the statistical relationship between link probability and distance (see, e.g., Liben-Nowell et al. (2005); Lambiotte et al. (2008); Goldenberg and Levy (2009); Krings et al. (2009) and the excellent reviews of Barthélemy (2011) and Kaltenbrunner et al. (2012)).

To the best of our knowledge, this paper is the first to propose a theory for the relationship between geographical distance and social interactions and to test it using the precise geometry of individual social contacts and the geographical distance between them. It is also the first that empirical establishes the degree of inefficiency of social interactions and, by using counterfactual exercises, determines whether it is more efficient to subsidize transportation costs or social interactions.

The rest of the paper unfolds as follows. Section 2 develops the theoretical model and determines the equilibrium while Section 3 studies its efficiency properties and the policy implications of the model. Section 4 is devoted to the empirical strategy. In Section 5, we describe our data, provide the empirical results and discuss them. In Section 6, we test the different predictions of the model and determine the level of inefficiencies of social interactions and social capital and how they are affected by the size of the network. We also simulate two policies and determine which one leads to the highest social welfare. Finally, Section 7 concludes the paper and discusses our policy results. All proofs in the theoretical

risk-sharing networks and Rosenthal and Strange (2008), Arzaghi and Henderson (2008), Bisztray, Koren, and Szeidl (2018) and List, Momeni, and Zenou (2019) who find that knowledge and productivity spillovers are important but decay sharply with distance.

[7] Another strand of related literature uses geographic proximity as a proxy for social interactions. Most notably, Bayer, Ross, and Topa (2008) assume that agents living in the same census block exchange information about jobs. Their finding that residing in the same block raises the probability of sharing work location by 33% is thus interpreted as a referral effect. Hellerstein, McInerney, and Neumark (2011); Hellerstein, Kutzbach, and Neumark (2014) and Schmutte (2015) build on the same assumption using matched employer-employee data with residential information. Using mobile phone data on one entire city in China, Barwick et al. (2019) show that geographical distance is important in spreading information about jobs.

model can be found in Appendix A. In Appendix B, we solve for the social capital fixed point and show under which condition it is unique. In Appendix C, we perform some robustness checks while we explain our calibration in the policy exercises in Appendix D.

2 The model

2.1 Notations and definitions

Consider a linear city on the line segment $x \in [-b, b]$ where b is the city border, and let $\lambda(x) : [-b, b] \to R^{+}$ measure the number of agents located at x. We focus on a city with unit mass population: $\int_{-b}^{b} \lambda(y)\mathrm{d}y = 1$.

Each agent *visits every other agent* and benefits from social interactions. First, the utility from social interactions is given by

$$S(x) = \int_{-b}^{b} v\left(n(x, y)\right) s(y)\lambda(y)\mathrm{d}y \tag{1}$$

where $n(x, y)$ is the number or, more exactly, the *frequency* of interactions that agent at x initiates with an agent at y who offers an interaction value $s(y)$.[8] For the sake of tractability, we assume that

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

This expression assumes decreasing returns to the frequency of interactions with a given agent; it even assumes negative returns (saturation) above $n = 1$. Observe that, in (1), we assume that there are decreasing returns in $v\left(n\left(x, y\right)\right)$ but not in $s(y)$. This is mainly for analytical tractability because we need to calculate a fixed point on social interaction and capital (see equations (9) and (10) below). Observe that Google's PageRank algorithm makes the same assumption when it computes the PageRank index.

Second, the interaction value offered by an agent residing at y is assumed to be equal to

$$s(y) = 1 + \alpha \int_{-b}^{b} n(y, z)s(z)\lambda(z)\mathrm{d}z \tag{3}$$

The first constant term (normalized to 1) represents the idiosyncratic interaction value that

[8] Here, as in Cabrales, Calvó-Armengol, and Zenou (2011), individuals do not explicitly choose with whom to link with but decide a level of social interactions at each location in the city.

the agent located at y provide to her visitors. The second term, $\alpha \int_{-b}^{b} n(y,z)s(z)\lambda(z)\mathrm{d}z$, reflects the value of her social network for her visitors. It increases with $n(y,z)$, the number of interactions, and $s(z)$, the value of her interactions. The parameter $\alpha > 0$ measures the importance of others' social capital in an agent's social capital formation. The higher is α, the higher is the impact of the social network of "friends of friends". We refer to $s(y)$ as the *social capital* of the agent located at y.

The social capital function $s(y)$ defined in (3) can be interpreted in various ways according to the context under discussion. In the context of information transmission (for example, about job opportunities) and/or knowledge (about a product or technique), the first term may represent the information endowed to or produced by the agent located at y while the second term may reflect the information she received during her visits to other agents. The parameter α then measures the imperfection of information transmission and its retention. In the context of a service sector like advertising, law, etc. (Arzaghi and Henderson, 2008), the first term represents the idiosyncratic productivity of a firm located at y while the second term reflects the potential and the ability to quickly subcontract parts of a project to other competent firms. In the context of friendship, community or political participation, the first term gives a measure of the pleasure or interest in a specific interaction (e.g., with a college friend, priest or politician) while the second term may reflect the sense of belonging to a community (e.g., alumni, confession or political group).

Third, each agent located at x incurs a cost of visiting another agent residing at y, $c(x-y)$, which is symmetric and increases with distance $|x-y|$: $c(z) = c(-z)$ and $c'(z) > 0 \; \forall z > 0$. For simplicity, we consider the class of travel cost functions $c(x)$ that are differentiable except at $x = 0$. We define the slope at $x = 0$ as $c'_{+}(0) \equiv \lim_{x \to 0, x > 0} c'(0) \geq 0$, recognizing the possible kink at $x = 0$. The total social interaction cost of an agent located at x is given by

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

which increases with the number of social interactions.

We now consider the question of how social capital is distributed across space when agents are exogenously located.

2.2 Social capital and space

We assume that λ, the population density at each location, is exogenously fixed. Each agent located at x chooses the profile of interactions $n(x,\cdot)$ that maximizes her utility

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

Note that her utility depends on the profile of other agent's social capital levels $(s(y), y \neq x)$. It also depends on her own social capital $(s(y), y = x)$ but only on a set of measure zero.[9] As a result, the optimal number of interactions of an agent located at x depends only on the social capital $s(y)$ of the other agents located at y at a non-zero distance to her. The optimal number of interactions $n^*(x,y)$ of an agent located at x (that we call agent x) is therefore found by *differentiating pointwise* $U(x)$ with respect to $n(x,y)$, taking $s(y)$ as given. This pointwise differentiation yields:

$$v'\left(n^*(x,y)\right) s(y) - c(x-y) = 0.$$

Using (2), this is equivalent to:

$$\left[1 - n(x,y)\right] s(y) = c(x-y).$$

So, the optimal number of interactions is equal to:

$$n^*(x,y) = 1 - \frac{c(x-y)}{s(y)} \tag{4}$$

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

$$n^*(x,y) > 0 \Leftrightarrow s(y) > c(x-y),\ \forall x, y$$

[9] Under the assumption that $\lambda(x) < +\infty$, the agent has no incentive to raise her number of interactions $n(x,\cdot)$ to increase her own social capital $s(x)$. In other words, since one agent's social capital benefits "almost" exclusively other agents, an agent has no incentives to be strategic with respect to increasing her own social capital.

A sufficient condition for this inequality to hold is

$$\min_y s(y) > c(2b) \tag{5}$$

Let us define the *access cost measure* as

$$g(y) \equiv \int_{-b}^{b} c(y-z)\lambda(z)\mathrm{d}z, \tag{6}$$

which is lower than the maximum travel cost $c(2b)$. By plugging (4) into (3) and using (6), we obtain the equilibrium level of social capital $s^*(y)$, which is given by:

$$s^*(y) = 1 + \alpha \int_{-b}^{b} s(z)\lambda(z)\mathrm{d}z - \alpha g(y). \tag{7}$$

Integral equations do not often accept simple analytical solutions, if any. Yet, under the above utility specification, a solution can be obtained. Indeed, integrating $s(z)\lambda(z)$ and simplifying, we obtain:

$$\int_{-b}^{b} s(z)\lambda(z)\mathrm{d}z = \frac{1}{1-\alpha}\left[1 - \alpha \int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z\right]. \tag{8}$$

Inserting this result into (7) yields a closed-form solution for the equilibrium social capital given by:

$$s^*(y) = s_0 - \alpha g(y), \tag{9}$$

where

$$s_0 = \frac{1 - \alpha^2 \int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z}{1-\alpha}, \tag{10}$$

and where $g(y)$ is defined by (6). Under the condition that $0 < \alpha < 1$, the optimal social capital $s^*(y)$ has a finite solution. To guarantee global interactions, we must have $s_0 - \alpha g(y) > c(x-y)$ for all x, y. Using (5), a sufficient condition is

$$s_0 - \alpha \left[\max_y g(y)\right] > c(2b) \tag{11}$$

To summarize,

Proposition 1 *Assume $0 < \alpha < 1$ and (11). Then, for all x, y, there exists a unique* *equilibrium* $(n^*(x,y), s^*(y))$ *such that*

$$n^*(x,y) = 1 - \frac{c(x-y)}{s^*(y)} \tag{12}$$

and

$$s^*(y) = \frac{1 - \alpha^2 \int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z}{1-\alpha} - \alpha \int_{-b}^{b} c(y-z)\lambda(z)\mathrm{d}z \tag{13}$$

Let us discuss the properties of the equilibrium social capital $s^*(y)$, defined in (13),[10] in a spatial environment.

First, lower travel costs increase social capital for all agents. This conclusion arises simply because social capital increases when the access measure $g(y)$ falls. An upward shift in the travel cost function $c(x)$ raises this access measure and therefore each agent's social capital $s^*(y)$. As a result, travel cost can be seen as a *barrier to social capital formation*. Improvements in urban transportation infrastructure should therefore enhance social capital.

Second, a rise in the importance of peers' social links in the creation of own social capital α, has ambiguous effects. Indeed, differentiating $s(y)$ with respect to α yields

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

where $s_\alpha(y)$ and $n^*_\alpha(y,z)$ denote the derivatives of $s(y)$ and $n^*(y,z)$ with respect to α. Thus, an agent's social capital increases with higher α because she places greater value on the social capital of her interaction partners (first term) and because her partners themselves have higher social capital (second term). However, as $n^*_\alpha(y,z) = -c(y-z)s^*_\alpha(z)/\left(s^*(z)\right)^2 \leq 0$, 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 (third term). We can get a clearer result by using the optimal frequency of interaction and its associated social capital (7). Differentiating the latter expression with respect to α leads to:

$$s_\alpha(y) = \int_{-b}^{b} s(z)\lambda(z)\mathrm{d}z - g(y) + \alpha \int_{-b}^{b} s_\alpha(z)\lambda(z)\mathrm{d}z. \tag{14}$$

[10] Once we know the comparative statics results with respect to $s^*(y)$, then it is straightforward to deduce those of $n^*(x,y)$.

Multiplying this expression by $\lambda(y)$, integrating and simplifying gives:

$$\int_{-b}^{b} s_\alpha(z)\lambda(z)\mathrm{d}z = \frac{1}{(1-\alpha)^2}\left[1 - \int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z\right]$$

Plugging this expression and (8) into (14) yields

$$s_\alpha(y) = \frac{1}{(1-\alpha)^2}\left[1 - \int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z\right] - g(y)$$

As expected, this expression is ambiguous in sign. However, it is positive for small enough access cost measure $g(\cdot)$ and therefore low enough travel costs $c(\cdot)$. We summarize these findings in the following proposition:

Proposition 2 *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.*

We now look at the impact on social capital of a *wider geographical dispersion of agents.* Consider a mean preserving increase in the spread of the spatial distribution λ; that is, a change in λ that *second-order stochastically dominates* the present distribution. Expanding expression (9), the social capital $s(y) = s_0 - \alpha g(y)$ can be found to be a linear function of

$$-\left((1-\alpha)\, g(y) + \alpha \int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z\right),$$

which can be rewritten as

$$-\int_{-b}^{b} \left[(1-\alpha)\, c(y-z) + \alpha g(z)\right] \lambda(z)\mathrm{d}z.$$

We can then apply standard results from the analysis of uncertainty. Namely, a mean preserving spread of λ will decrease this expression if the square-bracketed expression is a convex function of z for any y. Conversely, it will increase this expression if the square bracket term is a concave function of z for any y. A sufficient condition for a decrease (resp. an increase) of this expression is that both $c(\cdot)$ and $g(\cdot)$ are convex functions (resp. concave functions). For our class of travel cost functions, we find that

$$g''(x) = \int_{-b}^{b} c''(x-y)\lambda(y)\mathrm{d}y + 2c'_{+}(0)\lambda(x)$$

where $c'_{+}(0)$ is the positive slope at the possible kink of the travel cost function. Therefore $g(\cdot)$ is convex for any travel cost function that is piece-wise linear or convex. This includes linear travel cost $c(x) = c_1 |x|$ and quadratic travel cost $c(x) = c_2 x^2$ where c_1 and c_2 are constants. Intuitively, a spread of the spatial distribution of agents increases the trip distances and costs, which decreases the incentives to interact. So, *larger spatial dispersion of agents reduces social capital in cities.*[11]

Finally, agents located at the urban center have better access to others and have incentives to increase their social interactions and social capital. One therefore expects that social capital is less spatially dispersed than the agents. To make this argument formally, let us measure the *spatial dispersion* of a distribution function ϕ by the ratio of "spatial variance" over its mean value, i.e.,

$$\mathrm{Disp}(\phi) \equiv \frac{\int_{-b}^{b} z^2 \phi(z)\mathrm{d}z}{\int_{-b}^{b} \phi(z)\mathrm{d}z}$$

A mean preserving spread of the function ϕ around $x = 0$ increases this dispersion measure because it puts higher values to more distant locations. Under this definition, social capital is less spatially dispersed than the agents if and only if $\mathrm{Disp}(s\lambda) < \mathrm{Disp}(\lambda)$. Using (9), it is shown in the proof of Proposition 3 in Appendix A that this is equivalent to $\mathrm{Disp}(g\lambda) > \mathrm{Disp}(\lambda)$. That is, the function $g\lambda$ should be more dispersed than the agent's spatial distribution function λ. We further show that this is true irrespective of the travel cost function g when $x^2\lambda(x)/\int z^2\lambda(z)\mathrm{d}z$ is a mean preserving spread of the distribution of $\lambda(x)$ around its mean $x = 0$. This applies for any uniform spatial distribution λ and for most symmetric spatial distribution functions of interest. We summarize these results in the following proposition:

Proposition 3 *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;*

[11] Note that general results cannot be obtained for travel cost functions that are piece-wise concave (such as $c(x) = 1 - \exp(-|x|)$) because these functions are neither convex nor concave.

(*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* $x=0$.

The main point of Proposition 3 is to show that, 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. This implies that the level and the geographical dispersion of social capital are monotone functions of the dispersion of individuals. Even though the deteriorating effect of spatial dispersion on social interactions is not very surprising, we believe that this is the first theoretical proof (and empirical confirmation; see below) of this effect.

2.3 Linear travel costs

Let us now apply the above analysis to *linear travel costs*, which are heavily used in urban economics for their convenient and realistic properties (see, e.g., Fujita, 1989; Zenou, 2009). In the present paper, they permit closed-form solutions. Suppose, indeed, that $c(x) = c_1 |x|$ where $c_1 > 0$. Then,

$$\begin{aligned}
g(y) &\equiv c_1 \int_{-b}^{y} (y-z)\lambda(z)\mathrm{d}z + c_1 \int_{y}^{b} (z-y)\lambda(z)\mathrm{d}z \\
g'(y) &= c_1 \int_{-b}^{y} \lambda(z)\mathrm{d}z - c_1 \int_{y}^{b} \lambda(z)\mathrm{d}z \\
g''(y) &= 2c_1\lambda(y) > 0
\end{aligned}$$

So, the access cost measure g is a convex function of the distance to the center. Social capital is a concave function that is distributed so that $s''(y) = -2\alpha c_1\lambda(y) < 0$. Assume further that the spatial distribution of agents λ is *symmetric* ($\lambda(x) = \lambda(-x)$). Then, $g(x)$ is also symmetric and therefore equal to

$$g(x) = g_0 + 2c_1 \int_0^x \int_0^y \lambda(z)\mathrm{d}z\mathrm{d}y, \qquad x \geq 0$$

where $g_0 = 2c_1 \int_0^b z\lambda(z)\mathrm{d}z$. So, for $x \geq 0$, and assuming $0 < \alpha < 1$ and (11), then the unique equilibrium $(n^*(x,y), s^*(y))$ is given by

$$n^*(x,y) = 1 - \frac{c_1 |x-y|}{s^*(y)}$$

and

$$s^*(x) = s_0 - 2c_1 \int_0^x \int_0^y \lambda(z)\mathrm{d}z\mathrm{d}y$$

where

$$s_0 = \frac{1 - 2c_1\alpha^2 \left[\int_0^b z\lambda(z)\mathrm{d}z - 2\int_0^b \left(\int_0^x \int_0^y \lambda(z)\mathrm{d}z\mathrm{d}y\right)\lambda(x)\mathrm{d}x\right]}{1-\alpha}$$

It is clear that lower travel costs c_1 increase social capital for all agents. For small enough travel costs c_1, higher α increases s_0 and therefore each agent's social capital.

3 Efficient social interactions

We now study the planner's allocation of interaction frequency for a given location pattern λ. 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

$$s(x) \leq 1 + \alpha \int_{-b}^{b} n(x,z)s(z)\lambda(z)\mathrm{d}z \tag{15}$$

where we put an inequality to express that the agent can always reduce her social capital at no cost (e.g., she erases a part of her address book).

The government chooses the profiles $n(\cdot,\cdot)$ and $s(\cdot)$ that maximize the Lagrangian function

$$\mathcal{L} = \int_{-b}^{b}\int_{-b}^{b} \left\{ v\left[n(x,y)\right] s(y) - n(x,y)c(x-y) \right\} \lambda(y)\lambda(x)\mathrm{d}x\mathrm{d}y$$
$$- \int_{-b}^{b} \chi(x) \left[s(x) - 1 - \alpha \int_{-b}^{b} n(x,y)s(y)\lambda(y)\mathrm{d}y \right] \lambda(x)\mathrm{d}x$$

where $\chi(x) \geq 0$, or more precisely $\chi(x)\lambda(x)$ is the Kuhn-Tucker multiplier of the social capital constraint. So, $\chi(x)$ measures the welfare value of a marginal increase of the social capital of an agent located at x.

Lemma 4 *The efficient interaction frequency and social capital satisfy the following necessary conditions:*

$$v'\left[n(x,y)\right] s(y) - c(x-y) + \alpha\chi(x)s(y) = 0 \tag{16}$$

$$\int_{-b}^{b} \left\{ v\left[n(x,y)\right] + \alpha\chi(x)n(x,y) \right\} \lambda(x)\mathrm{d}x - \chi(y) = 0 \tag{17}$$

Equations (16) and (17) together with the constraint (15), solve for the functions $n(x,y)$, $s(y)$ and $\chi(x)$.

Condition (16) captures the main externality at work in the process of social interaction. When the planner chooses the interaction frequency $n(x,y)$, he considers both the benefit and cost to agent x and the fact that an increase in x's social capital increases y's social capital. This latter effect is not considered by agent x at the 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, $\chi(x)$.

The second condition (17) is interpreted as follows: when the planner increases the social capital of an agent located at y, he directly raises the utility of all agents who interact with this agent (first term in curly brackets) and indirectly increases the social capital for all those other agents (second term in the curly brackets). In the efficient allocation, this combined effect should be equal to $\chi(y)$, the welfare value of a marginal increase of the social capital of an agent located at y.

Proposition 5 *The equilibrium frequency of interactions and level of social capital are lower than the efficient ones.*

Intuitively, the planner internalizes the effect that each agent has on others' social capital when she entertains more intense social interactions. As a result, the planner imposes agents to increase their frequency of social interactions above the equilibrium level. This welfare conclusion confirms Brueckner and Largey's (2008) and extends their analysis to the case where agents are distributed across space.

Can the efficient allocation of social interactions be decentralized with subsidies $\sigma(x,y)$ and $\tau(x,y)$ for social interactions and travel costs? If we include these subsidies, the utility becomes

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

This implies that the equilibrium number of social interactions becomes

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

We can obtain the first-best solutions and efficient social interactions can therefore be decentralized by setting $\sigma(x,y)=0$ and $\tau(x,y)=\alpha\chi^o(x)s^o(y)$. Indeed, in this case, we find: $n^*(x,y) = 1 - c(x-y)/s^o(y) + \alpha\chi^o(x) = n^o(x,y)$.

Proposition 6 *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.*

The optimal subsidy to travel costs is therefore not a uniform one. This suggests that decentralization would be difficult to implement because subsidies depend on both the origins and destinations of social interactions (it is very unlikely that $\tau(x,y)$ reduces to a simple function of x, or y or $x-y$). This result contrasts with Helsley and Zenou (2014), who advocate that the planner should subsidize the most central agents. Their model with a two location points, however, imperfectly captures the full picture of spatial interactions. In the present model, we observe that the planner does not subsidize those agents with high social capital but only subsidizes the trips to those agents.

4 Empirical strategy

To bring the model to the data, we need to introduce agents' heterogeneity in equation (2). 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.

Pointwise differentiating the utility

$$U(x)=S(x)-C(x)=\int_{-b}^{b}\left\{v\left(n(x,y)\right)s(y)-n(x,y)c(x-y)\right\}\lambda(y)\mathrm{d}y \tag{18}$$

with respect to $n(x,y)$, we easily obtain the optimal number of interactions, which is equal to:

$$n^*\left(x,y\right)=n_0-\frac{c\left|x-y\right|}{s^*(y)}+\theta(x,y),$$

while the social capital of each individual is still given by (3), which is equal to:

$$s(y)=1+\alpha\int_{-b}^{b}n(y,z)s(z)\lambda(z)\mathrm{d}z \tag{19}$$

Let us assume that we observe data from R networks ($r=1,...,R$), each having N_r agents. To avoid cumbersome notation, we assume that 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 $d_{ij,r}$. As a result, the above two equations can be written as follows:[12]

$$n^*_{ij,r}=n_0-\frac{cd_{ij,r}}{s^*_{j,r}}+\theta_{ij,r}, \tag{20}$$

and

$$s^*_{j,r}=1+\alpha\sum_{k=1}^{N_r}n^*_{jk,r}s^*_{k,r}, \tag{21}$$

Observe that, quite naturally, we do not allow social interactions with oneself, i.e., we assume $n^*_{ii,r}=0$.

[12] Observe that, for the purpose of the empirical analysis, (21) is a discrete version of (3).

We allow the social distance to depend on observed (pair-level) individual characteristics $x_{ij,r}$ and on unobserved factors $\varepsilon_{ij,r}$. For simplicity, we assume that $\varepsilon_{ij,r}$ is independent and identically distributed across pairs and networks, but the i.i.d. assumption within a network can be relaxed.

If the network is *undirectional,* that is if $n_{ji,r} = n_{ij,r}$ for all i, j, one can use the specification:

$$\theta_{ij,r} = \sum_{m=1}^{M} \beta_m |x_{i,m,r} - x_{j,m,r}| + \sum_{m=1}^{M} \beta_{M+m}(x_{i,m,r} + x_{j,m,r}) + \varepsilon_{ij,r}, \tag{22}$$

where negative values in the vector $(\beta_1, \ldots, \beta_M)$ capture *homophily* effects (associated with smaller socio-economic distance $|x_{i,m,r} - x_{j,m,r}|$), and $(\beta_{M+1}, \ldots, \beta_{2M})$ measures the effect of the combined level of x_i and x_j, where M is the number of individual-level covariates. Indeed, under homophily behavior (i.e., the tendency of individuals to associate and bond with others who share common traits; see McPherson, Smith-Lovin, and Cook, 2001; Currarini, Jackson, and Pin, 2009; Graham, 2017), individuals with similar characteristics (same race, same gender, etc.) will tend to interact more than less similar individuals (thus β_m should be negative under homophily in x_m).

If the network is *directional,* that is when $n_{ij,r}$ does not need to be equal $n_{ji,r}$, one can use the specification:

$$\theta_{ij,r} = \sum_{m=1}^{M} \beta_m (x_{i,m,r} - x_{j,m,r}) + \sum_{m=1}^{M} \beta_{M+m}(x_{i,m,r} + x_{j,m,r}) + \varepsilon_{ij,r} \tag{23}$$

Similar specifications have been used in the literature; see, for example, Fafchamps and Gubert (2007).

By plugging the value of $n^*_{ij,r}$ from (20) into (21), in Appendix B, we solve for the social capital fixed point and show under which condition it is unique. The social capital fixed point is given by:

$$\mathbf{s}^*_r = \left[\mathbf{I}_{N_r} - \alpha\left(\mathbf{N}_0 + \boldsymbol{\Theta}_r\right)\right]^{-1}\left(\mathbf{I}_{N_r} - \alpha c \mathbf{D}_r\right)\mathbf{1}_{N_r}, \tag{24}$$

where $\mathbf{s}_r$ is the $(N_r \times 1)$ vector with elements $s_{i,r}$, $\mathbf{I}_{N_r}$ is the $(N_r \times N_r)$ identity matrix, $\mathbf{1}_{N_r}$ is the $(N_r \times 1)$ vector of 1, and $\mathbf{N}_{0,r}$ is an $(N_r \times N_r)$ matrix in which the off-diagonal elements are n_0, and the diagonal elements are zero, while $\mathbf{D}_r = (d_{ij,r})$ and $\boldsymbol{\Theta}_r = (\theta_{ij,r}) = (x^{\mathrm{T}}_{ij,r}\beta + \varepsilon_{ij,r})$ are $(N_r \times N_r)$ matrices (see (B.6) in Appendix B.).

Estimation strategy For each network r, our dataset provides us with $x_{ij,r}$, the agents' characteristics, $n^*_{ij,r}$, the intensity of social interactions between agents i and j, $d_{ij,r}$, the geographical distance between agents i and j, and N_r, the number of agents in the network. Using this information, we will recover the structural parameters α, β, c, n_0, and the equilibrium social capital, $s^*_{j,r}$. For that, we employ the indirect inference (I-I) estimation method, proposed by Gourieroux, Monfort, and Renault (1993), that recovers the true parameters from the data by attempting to closely match simulated and observed levels of social interactions. The estimator is indirect in the sense that, rather than directly estimating the structural model, it estimates an *auxiliary* model that can be estimated with (computationally) easier methods such as the ordinary least squares (OLS). We run the auxiliary model with the observed data and the simulated ones. The estimates for the structural parameters are ones that best match the two sets of auxiliary parameters, based on an injectivity assumption (i.e., one-to-one mapping between the structural parameters and the auxiliary parameters).

Structural model For the sake of exposition, we denote the vector of structural parameters by $\mu \equiv (n_0, \alpha, c, \beta)$, we group the unobserved information into the vector $\mathcal{E}_r \equiv (\varepsilon_{ij,r})$ and the observed information into the vector $\mathbf{Y}_r \equiv (\mathbf{X}_r, \mathbf{D}_r, N_r)$ where $\mathbf{X}_r$ and $\mathbf{D}_r$ capture the individuals characteristics $x_{i,r}$ and the distances $d_{ij,r}$, respectively. The structural model (20) and (24) can now be written as the following system of equations:

$$n^*_{ij,r}(\mathbf{Y}_r, \mathcal{E}_r; \mu) = n_0 - \frac{c d_{ij,r}}{s^*_j(\mathbf{Y}_r, \mathcal{E}_r; \mu)} + x_{ij}^{\mathrm{T}} \beta + \varepsilon_{ij,r}, \tag{25}$$

$$\mathbf{s}^*(\mathbf{Y}_r, \mathcal{E}_r; \mu) \equiv \left[\mathbf{I}_r - \alpha \left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)\right]^{-1} \left(\mathbf{I}_r - \alpha c \mathbf{D}_r\right) \mathbf{1}_r, \tag{26}$$

Auxiliary model We use simple linear regression equations as an auxiliary model. We propose a first auxiliary model equation that expresses the relationship between social interaction intensities, individual characteristics and distance between interaction partners as follows:

$$n_{ij,r} = \gamma_{10} + x_{ij,r}^{\mathrm{T}} \gamma_{11} + \gamma_{12} d_{ij,r} + \epsilon_{1,ij,r}. \tag{27}$$

We propose a second auxiliary model equation expressing a similar relationship with respect to indirect interactions. Let us denote by $\mathbf{N}_r$ the matrix of social interaction intensities for network r, where its ith row and jth column element is $n_{ij,r}$. We further define the matrix of second degree interaction as the square matrix $\mathbf{N}_r^2 \equiv \mathbf{N}_r \mathbf{N}_r$. We denote by $[\mathbf{N}_r^2]_{ij}$ the ith row and jth column element of this matrix. The second auxiliary equation can then be written as:

$$[\mathbf{N}_r^2]_{ij} = \gamma_{20} + x_{ij,r}^{\mathrm{T}} \gamma_{21} + \gamma_{22} d_{ij,r} + \epsilon_{2,ij,r}. \tag{28}$$

We denote by $\boldsymbol{\gamma}$ the vector of the above auxiliary model coefficients.

Algorithm We draw T sets of simulation errors, $\mathcal{E}^t \equiv (\varepsilon_{ij,r}^t)$, $t = 1, \ldots, T$, for all pairs i and j and all networks r. These sets of errors are fixed for the entire estimation process.[13] First, we compute social capital $\mathbf{s}_r^t$ and predict the intensity of social interactions $\widehat{n}_{ij,r}^t$ for each set of errors using equations (25) and (26). To match the data, we constrain $\widehat{n}_{ij,r}^t$ to lie between zero and five (included). This process yields the first degree interaction matrix $\widehat{\mathbf{N}}_r(\mathcal{E}_r^t, \mathbf{Y_r}; \mu)$ and the second degree interaction matrix as the square of the latter. Let $\mathbf{Y}$, $\mathcal{E}^t$, $\mathbf{N}$ and $\widehat{\mathbf{N}}(\mathcal{E}^t, \mathbf{Y}; \mu)$ collect the observed data, the non-observed data, the observed interactions and the predicted interactions in all networks. We then run OLS regressions on the auxiliary model (27) and (28) separately with the observed and simulated interaction values. As a result, we obtain a set of the OLS estimates $\widehat{\boldsymbol{\gamma}}(\mathbf{N}, \mathbf{Y})$ with the observed interactions and a set of estimates $\widehat{\boldsymbol{\gamma}}[\widehat{\mathbf{N}}(\mathcal{E}^t, \mathbf{Y}; \mu), \mathbf{Y}]$, $t = 1, \ldots, T$ with the simulated interactions. Finally, since OLS estimates using the simulated data are functions of the structural parameter vector μ, we choose μ that leads the closest difference between $\widehat{\boldsymbol{\gamma}}(\mathbf{N}, \mathbf{Y})$ and $\widehat{\boldsymbol{\gamma}}[\widehat{\mathbf{N}}(\mathcal{E}^t, \mathbf{Y}; \mu), \mathbf{Y}]$. Formally, the I-I estimator $\widehat{\mu}_{\mathrm{II}}$ is constructed such that

$$\widehat{\mu}_{\mathrm{II}} = \arg\min_{\mu} \left|\left| \widehat{\boldsymbol{\gamma}}(\mathbf{N}, \mathbf{Y}) - \frac{1}{T} \sum_{t=1}^{T} \widehat{\boldsymbol{\gamma}}[\widehat{\mathbf{N}}(\mathcal{E}^t, \mathbf{Y}; \mu), \mathbf{Y}] \right|\right|, \tag{29}$$

where the norm $|| \cdot ||$ is defined by a (positive-definite) weight matrix, $\mathbf{A}$, with dimension equal to the number of the auxiliary model parameters. Gourieroux, Monfort, and Renault (1993) show that the efficient weight matrix is given by the inverse of the variance of the moment conditions in (29), evaluated at the true parameter value μ_0. Hence, we use

$$\mathbf{A} = \left[\left(1 + \frac{1}{T}\right) var(\widehat{\boldsymbol{\gamma}}(\mathbf{N}, \mathbf{Y})) \right]^{-1} \tag{30}$$

[13] Gourieroux, Monfort, and Renault (1993) show that the I-I estimator is consistent for a fixed number of simulation draws.

as our weight matrix. We estimate **A** using a bootstrap (e.g., Ackerberg and Gowrisankaran, 2006). Given the complex dependence structure of dyadic observations within each network, we also use a bootstrap to calculate the standard errors of our estimated structural parameters, where we resample networks instead of individuals to address clustering at the network level.

Identification Our model consists of four main parameters: the baseline social interaction intensity n_0, the social capital accumulation parameter α, the transportation cost c, and the effect of social distance β. Understanding the separate identification of each of these parameters is challenging because our model is nonlinear and our error terms are not additively separable, which is more complicated than a typical model of network externalities, such as the linear-in-means network model (Manski, 1993). Although matching the OLS estimates of the auxiliary model between the observed and simulated data yields reasonable estimates of the structural parameters, it is important to discuss the identification of the model.

To illustrate the separate identification of these four parameters more formally, we focus on the sources of identification. First, consider β. In the first equation (27) of the auxiliary model, it is straightforward to assume that there is a one-to-one relationship between γ_{11} and β, as equation (27) closely mimics equation (25) in x_{ij} term. The intercept, or the baseline intensity level, n_0, is similarly identified from its one-to-one relationship with γ_{10}. Next, the cost parameter c is identified given that both equations (27) and (28) contain the term $d_{ij,r}/N_r$. Given that the cost parameter is a coefficient on $d_{ij,r}/s_j^*$ in equation (25), having γ_{22} in addition to γ_{12} helps the identification of c.

The most challenging (structural) parameter to identify is α in (26). To obtain α, consider the social capital equation (21).

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

Social capital is recursively defined, and hence, it is a function of not only the first degree network connections (or social interactions) but also further-degree indirect connections. Therefore, we use the additional equation (28), which uses $[\mathbf{N}_r^2]_{ij}$, the number of second-degree interactions between i and j as a dependent variable, to identify the importance of others' social capital in an agent's social capital formation. The overall fit of two auxiliary equations, measured by R^2 will help the identification of the social capital parameter α. Since the identification of α may depend on more than second-degree social interactions, we do a robustness check by including additional equation for the third-degree interactions in Appendix C.

5 Empirical analysis

5.1 Data

Our empirical investigation is made possible by the use of a database on friendship networks from the National Longitudinal Survey of Adolescent Health (Add Health).[14]

The Add Health database has been designed to study the impact of the social environment (i.e., friends, family, neighborhood and school) on adolescents' behavior in the United States. It is a school-based survey which contains extensive information on a representative sample of students who were in in grades 7–12 in 1995. More than 100 schools were sampled. Three features of the Add Health data set are unique and central to our analysis: *(i)* the nomination-based friendship information, which allows us to reconstruct the precise geometry of social contacts, *(ii)* the detailed information about the intensity of social interactions between each of two friends in the network; and *(iii)* the geo-coded information on residential locations, which allows us to measure the geographical distance between individuals.

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).[15] For each individual i, the friendship nomination file also contains detailed information on the frequency and nature of interaction with each nominated friend j. The precise questions are: "Did you go to {NAME}'s house during the past seven days?"; "Did

[14] This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

[15] The limit in the number of nominations is not binding (even by gender). Less than 1% of the students in our sample show a list of ten best friends.

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?". "Yes" or "No" are the possible answers. These answers are coded by one and zero, respectively. We measure the intensity of social interactions between students i and j, that is $n(x, y)$ or n_{ij} in the model, by summing all these items so that the maximum value of n_{ij} is 5 and the minimum is 0. When we use the symmetric model, we choose the largest number between n_{ij} and n_{ji}. The average intensity of social interactions is 1.05, with noticeable dispersion around this mean value (standard deviation equal to 1.64). A random sample of students in each school (about 20,000 students) is then interviewed also at home where a longer list of questions are asked both to the child and his/her parents. Most notably for this study, the geographical locations of those houses is also 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 by computing the Euclidean distance between their homes. The maximum geographical distance between two students in a network is about 47 kilometers. The average distance is 6.75 kilometers, and its standard deviation is 6.71 kilometers.

When data on individuals with geo-coded information are merged with the friendship nomination data, valid information on nominated friends, types of interactions and geographical location is available for 4,439 students.[16] Then, we focus on network sizes between 4 and 10 members for the following two reasons. Firstly, the upper and lower tails of the distribution of networks by network size are commonly trimmed since the strength of peer effects may be too different in too small or too large networks (see Calvó-Armengol, Patacchini, and Zenou, 2009). Secondly, and most importantly, to reduce the discrepancy between the theoretical model, which assumes that all individuals interact with each other (i.e., $n_{ij} > 0$), and the data, for which many students are not friends with each other (i.e., $n_{ij} = 0$), we keep the sample size of the networks relatively small, so that students are more likely to be friends with each other.[17] Observe that, if, as in Section 2.1, we interpret λ as a friendship probability, then, there is no inconsistency between the theory and the data since, in this

[16] A large reduction in sample size when mapping friendships in the Add Health is common and mainly due to the network construction procedure – roughly 20 percent of the students do not nominate any friends and another 20 percent cannot be correctly linked. In addition, there is a further 50 percent of the sample for which information on strength of interactions is missing.

[17] In Appendix C, we check the robustness of our estimation results to the sample-size reduction.

interpretation, each student is not friend with every other student, as it is the case in the data, where students are shown to have few contacts.

Our final sample consists of 753 individuals distributed over 141 networks. Table 1 describes our data and details our sample selection procedure. We report the characteristics of four different samples, which correspond to the three steps of our selection procedure. In column (1), we consider the original sample of students who have valid geo-coded information. In columns (2)–(3), we further restrict the sample to those with friendship information and intensity of interactions. Finally, in column (4), we report our sample where we only keep students in networks of 4–10 agents.

Table 1 also shows that differences in means between these samples are almost never statistically significant, which strongly suggests no specific bias in the selection of the sample. Among the adolescents selected in our sample of students, 53% are female and 18% 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 2.86, meaning slightly less than a grade of “B.” The average family income is 48,410 in 1994 dollars, although 10% of parents chose not to report such information. The average number of social interactions is 1.05.

[*Insert Table 1 here*]

5.2 Empirical results

Table 2 displays the estimation results. In column (1), we report the estimations of equation (23), i.e. when the dyadic relationship is *directional* implying $(x_i - x_j)$. In column (2), we display the estimation results of (22), i.e., when the dyadic relationship is *undirectional* implying $|x_i - x_j|$. We include socio-demographic characteristics possibly related to the intensity of social interactions and social capital, such as household size. The estimates are remarkably similar across columns (1) and (2). Moreover, the estimates for the main structural parameters, n_0, α, c, do not differ substantially across the two specifications.

[*Insert Table 2 here*]

Table 2 gives information about the effects of individual characteristics on interactions.

When we consider the symmetric social distances, i.e., column (2), students' preferences exhibit homophily in their own characteristics if the coefficient β_m is negative and significantly different from zero. This occurs for most individual characteristics: female, ethnicity, grade, GPA, physical development, religion practice, and whether they refuse to answer family income. The estimates are all negative and significant, which also supports homophily behaviors. The magnitudes of the estimates are in general larger than those from the directed social distances in column (1). When it comes to family background, we find strong homophily behaviors in family income, having two parents, and family size, but no homophily in parental education. The degree of homophily is the largest in gender.

Next, the estimated coefficients on the $(x_i + x_j)$ variables exhibit mixed signs. The results in columns (1) and (2) are mostly similar to each other. In both specifications, the intensity of social interactions is increasing if a pair of students are older (i.e., a higher grade), male, and black students, they are more religiously practicing, or they have more family income, in terms of total endowment of both students in a pair. By contrast, the intensity of social interactions is decreasing if students have a higher GPA, or they are from families with two parents, more family members, or more income.

Turning our attention to the structural parameters, we see that they are all statistically significant and have reasonable values. Indeed, the estimated baseline level of social interactions n_0 is roughly between 1.01 and 1.40, which implies a pair of students want to have a social interaction if they have the same characteristics (and the zero level of combined characteristics). The estimated cost of transportation c is 0.19–0.22 across specifications. After we multiply the cost with the average pairwise distance (6.71 kilometers), the average estimated transportation cost is 1.25–1.48. Given the estimated cost of transportation per kilometer, the magnitudes of homophily parameters (i.e., the coefficient estimates on $|x_i - x_j|$) ranging between 0.02 and 0.44 are not small. Finally, α, which measures the importance of others' social capital in an agent's social capital formation, has an estimated value of 0.08, which implies possible positive externalities. This implies that there may be inefficiency in the levels of equilibrium social interactions if students do not consider such externalities. We will investigate this further in our policy simulations.

6 Policy analysis

6.1 Welfare

We now use the estimated parameters of the model provided in Table 2, i.e., α, c and n_0, to calculate the welfare loss and to perform some simulations. We know from the theoretical model (Section 3) that, if the planner optimally chooses $n(x,y)$ and $s(y)$, we obtain:

$$v'\left[n^o(x,y)\right]s^o(y) - c(x-y) + \alpha\chi(x)s^o(y) = 0$$

$$\int_{-b}^{b}\left\{v\left[n^o(x,y)\right] + \alpha\chi(x)n^o(x,y)\right\}\lambda(x)\mathrm{d}x - \chi(y) = 0$$

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

where the Kuhn-Tucker multiplier $\chi(x) \geq 0$ measures the welfare value of a marginal increase of the social capital of an agent located at x. In this welfare analysis, we assume linear travel cost and discretize the above conditions as

$$n^o_{ij,r} = n_0 - \frac{cd_{ij,r}}{s^o_{j,r}} + \alpha\chi_{i,r}\, s^o_{j,r} + \theta_{ij,r}, \tag{31}$$

$$\chi_{j,r} = \sum_{i=1}^{N_r}\left\{\left(n_0 + \theta_{ij,r}\right)n^o_{ij,r} - \frac{1}{2}\left(n^o_{ij,r}\right)^2 + \alpha\chi_{i,r}\, n^o_{ij,r}\right\}, \tag{32}$$

$$s^o_{j,r} = 1 + \alpha\sum_{k=1}^{N} n^o_{jk,r}s^o_{k,r}. \tag{33}$$

Here is how we proceed. From the previous estimations of the equilibrium model, we have the estimated values of n_0, α, c and $\theta_{ij,r}$ (Table 2). From the data, we know $d_{ij,r}$. By plugging these values into (31), (32) and (33), 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 $\chi_{i,r}$ for all i. For each network r, we have $2N_r + L_r$ unknowns, where L_r is the number of links in network r, and we have $2N_r + L_r$ equations since there are L_r equations for (31), N_r equations for (32) and N_r equations for (33). We then compare the observed equilibrium values of $n^*_{ij,r}$ and $s^*_{j,r}$ with the social optimum values $n^o_{ij,r}$ and $s^o_{j,r}$ (using equations (26) and (33) evaluated at our parameter estimates). According to Proposition 5, we should find that students socially interact too little compared to the social optimal outcome, i.e., $n^o_{ij,r} > n^*_{ij,r}$, $\forall i, j$, and $s^o_{i,r} > s^*_{i,r}$, $\forall i$.

We numerically solve the optimal level of social interactions and social capital with the I-I parameter estimates displayed in column (2) in Table 2 by running a total of 100 simulations. Table 3 displays the results. Note that, in this table, we first take the average of social interactions in each network and, then, take the average again over all networks. We find that each pair interacts on average 1.89 fewer times than is socially optimal. The difference between the socially optimal and the observed levels of social interactions varies from 0.53 to 3.76 across networks. Although there are a few networks where the observed level is larger than the optimal level, many networks' interactions fall short of the optimum. Students also have less social capital than optimal (by 0.83, or approximately 34%, on average).

[*Insert Table 3 here*]

Network size and social interactions We would now like to find which variables are closely associated with the discrepancy between the optimal level and the observed level. For that, we regress the differences $\overline{n}^o_r - \overline{n}^*_r$ and $\overline{s}^o_r - \overline{s}^*_r$ on the network size, network measures, and average characteristics (e.g., average family income) of students in each network r:

$$\overline{n}^o_r - \overline{n}^*_r = \gamma_0 + \gamma_1 N_r + \gamma_2 \left(N_r\right)^2 + \gamma_3 \overline{d}_r + \gamma_z z_r + \gamma_x x_r + \epsilon_r, \tag{34}$$

$$\overline{s}^o_r - \overline{s}^*_r = \delta_0 + \delta_1 N_r + \delta_2 \left(N_r\right)^2 + \delta_3 \overline{d}_r + \delta_z z_r + \delta_x x_r + \zeta_r. \tag{35}$$

Tables 4–5 display the results. Consider, first, *social interactions* (Table 4) and let us examine if the difference between the optimal and the observed level of social interactions, $(\overline{n}^o_r - \overline{n}^*_r)$, is increasing or decreasing with network size N_r. Using column (5), we have:

$$\frac{\partial(\overline{n}^o_r - \overline{n}^*_r)}{\partial N_r} = \gamma_1 + 2\gamma_2 N_r = 1.502 - 2(0.089)N_r = 0 \tag{36}$$

Solving this equation leads to: $N_r = \frac{1.502}{2(0.089)} = 8.44$. This means that the difference between the optimal and the observed level of social interactions is increasing until the network size reaches (approximately) 8 students and then decreases. As a result, there is a non-monotonic relationship between $\overline{n}^o_r - \overline{n}^*_r$ and N_r where an increase in the network size increases $\overline{n}^o_r - \overline{n}^*_r$ up to $N_r = 8$ and, above this size, an increase in the network size decreases $\overline{n}^o_r - \overline{n}^*_r$. Thus,

$N_r = 8$ is the size of the network that *maximizes* these inefficiencies. Given that the median size of the networks is $N_r^{med} = 5$, in terms of magnitude, an increase by one person in the network from $N_r^{med} = 5$, raises these inefficiencies by $1.502 - 2(0.089)N_r^{med} = 0.61$.

[*Insert Tables 4 and 5 here*]

The average geographic distance is not significantly associated with the inefficiency. A one-kilometer increase in the average pairwise distance lead to a 0.0025 decrease in the inefficiency, but the magnitude is small and insignificant. Only a few average characteristics of the students are associated with the optimal-observed difference in social interactions. In particular, networks that consist of students with a higher average grade (and hence age) or a low GPA are more likely to have low inefficiencies in terms of social interactions.

Let us now turn to the inefficiencies in terms of social capital (Table 5). We find that the network population and the average geographic distance are not strongly associated with the inefficiency in social capital.

Although these regressions do not have a formal identification strategy, the results, partly based on the structural estimation of the model (that determine $\overline{n}_r^o - \overline{n}_r^*$ and $\overline{s}_r^o - \overline{s}_r^*$), provide some interesting explanations on what drives the size of inefficiency of the intensity of social interactions and social capital accumulation.

Network size and average welfare Another interesting exercise, for which we do not have a theory, is to determine the optimal network, i.e., the one that maximizes total welfare.[18] For that, without any policy, we compare the average welfare (to avoid size effects, the welfare is not defined as the sum of utilities but as the average utility) in each of the 141 networks. Remember that the welfare in network r is given by:

$$W_r^* = \sum_{i=1}^{N_r} \sum_{j=1}^{N_r} \left[\left((n_0 + \theta_{ij,r})\, n_{ij,r}^* - \frac{1}{2} \left(n_{ij,r}^*\right)^2 \right) s_{j,r}^* - n_{ij,r}^* c d_{ij,r} \right] \tag{37}$$

[18] Determining the optimal network is a very difficult exercise; see König, Tessone, and Zenou (2014) and Belhaj, Bervoets, and Deroïan (2016) for such attempts when the network is given. Jackson and Wolinsky (1996) provide a similar exercise for endogenous network formation. Because this exercise is complicated, only extreme structures emerge such as the complete network, the star network or nested split graphs. This is why we do it here by numerical simulations based on the estimated parameters.

As a result, the average welfare per network is:

$$AW_r^* = \frac{W_r^*}{N_r}$$

We would like know which network size N_r yields the largest AW_r^*.

For that, we run the following regression:

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

to investigate the relationship between average welfare and network size. In addition, as controls, we include the average geographical distance and network measures, such as mean and standard deviation of the degree distribution, average eigenvector centrality, clustering coefficient, and diameter.[19] We include the network measures (such as average degree and average eigenvector centrality of a network) to see how the shape of a network is associated with the welfare.

Table 6 reports the results. We can first calculate the network size that maximizes the average welfare per network AW_r^*. Using column (5), we have:

$$\frac{\partial AW_r^*}{\partial N_r} = \delta_1 + 2\delta_2 N_r = 20.87 - 2(0.939)N_r = 0 \tag{38}$$

Solving this equation leads to: $N_r = \frac{20.87}{2(0.939)} \approx 11$. This means the network that comprises (approximately) 11 students is the one that maximizes the average welfare per network. This is a network of a relatively large size. Since our sample include networks of max 10 students, we can only state that the size of all our networks is too small compared to the one that maximizes average welfare.

[*Insert Table 6 here*]

In Table 6, we also find that the average pairwise geographic distance is an important factor for designing an optimal network. The higher is the distance between two individuals in a network, the lower is the average welfare. In addition, from the changes in R^2 across columns (1) and (2), from 0.11 to 0.55, we find that average geographic distance explains a significant portion of the average welfare in a network.

[19] We compute the clustering coefficient as the ratio of the number of triangle loops to the number of connected triples.

6.2 Policies

We have seen in Proposition 6 that the social optimal allocation can be restored if social interactions are not subsidized while commuting trips are subsidized as a function of the locations of the destination and origin partners. Because the latter policy requires detailed information about every interaction pair, it is unlikely to be implemented. In this section, we consider the more realistic case of uniform subsidies on social interactions and/or travel costs that only target each individual irrespective of their personal characteristics but not a pair of individuals. We evaluate their impact on the frequency of interactions, $n(x, y)$ by running a total of 100 policy simulations. Which policy is more effective at moving the observed interactions/social capital closer to the optimal levels?

Assume that each individual receives a fixed subsidy σ for each interaction made with a friend and a (percentage) subsidy τ on her transport cost c. The total amount of each subsidy received by an individual located at x is therefore given by $\int_{-b}^{b} \sigma n(x, y)\lambda(y)\mathrm{d}y$ for social interactions and $\int_{-b}^{b} n(x, y)\tau c\, |x - y|\, \lambda(y)\mathrm{d}y$ for transportation costs. Note that the government (or the planner) is here introduced as an agent that can set subsidies on social-interaction efforts before the individuals decide upon their efforts. The assumption that the government can pre-commit itself to such subsidies and thus can act in this leadership role is fairly natural. As a result, this subsidy will affect the levels of social interaction efforts of all individuals.[20]

For each individual located at x and interacting with someone at y, when subsidies are included, the first-order conditions lead to the following level of social interactions

$$n^*(x, y) = \left[n_0 - \frac{c\,|x - y|}{s^*(y)} + \theta(x, y) \right] + \frac{\sigma}{s^*(y)} + \frac{\tau c\,|x - y|}{s^*(y)},$$

while the social capital is still given by

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

Holding social capital constant, quite naturally, the subsidies increase the number of social interactions. Subsidies can entice interactions with new partners as the number of interactions to a partner may rise from zero to a positive value in the presence of the subsidy. The

[20] This is is similar to the standard policy of firms' subsidies on R&D efforts; see e.g., Spencer and Brander (1983) and König, Liu, and Zenou (2019).

total welfare is now defined as:[21]

$$
\begin{aligned}
W = & \int_{-b}^{b}\int_{-b}^{b}\left[\left(\left(n_0+\theta(x,y)\right)n^*\left(x,y\right)-\frac{1}{2}\left[n^*\left(x,y\right)\right]^2\right)s^*(y)-n^*(x,y)c\left|x-y\right|\right]\lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y \\
& +\int_{-b}^{b}\int_{-b}^{b}n^*(x,y)(\sigma+\tau c\left|x-y\right|)]\lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y.
\end{aligned}
$$

6.2.1 Subsidizing social interactions

We consider a uniform subsidy σ_r for each network. We use the following discrete version of the equilibrium identities:

$$n_{ij,r}^{\sigma}=n_0+\frac{\sigma_r-cd_{ij,r}}{s_{j,r}^{\sigma}}+\theta_{ij,r} \tag{39}$$

and

$$s_{j,r}^{\sigma}=1+\alpha\sum_{k=1}^{N_r}n_{jk,r}^{\sigma}\,s_{k,r}^{\sigma} \tag{40}$$

where the superscript σ denotes the subsidy policy outcome. For the estimation, the total welfare per network is equal to

$$W_r^{\sigma}=\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\left[\left(\left(n_0+\theta_{ij,r}\right)n_{ij,r}^{\sigma}-\frac{1}{2}\left(n_{ij,r}^{\sigma}\right)^2\right)s_{j,r}^{\sigma}-\left(cd_{ij,r}-\sigma_r\right)n_{ij,r}^{\sigma}\right] \tag{41}$$

In this exercise, we determine the subsidy σ_r^* that gives network r the same aggregate welfare W_r^{σ} as its first best level 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 N_r. We then numerically solve equations (39) and (40) and find the subsidy such that $W_r^{\sigma}=W_r^{0}$. See Appendix D for technical details.

The first three columns in Table 7 display the results. On average, a subsidy level of 0.826 for each social interaction is required for a network to achieve the first-best aggregate level of social interactions and social capital.

[*Insert Table 7 here*]

[21] Observe that the optimal subsidy algorithm that we used is designed to find a maximum, so the optimal subsidy found corresponds to the (local) maximum.

6.2.2 Subsidizing transportation costs

In the case of subsidies on transport cost, we consider the following discrete versions of the equilibrium conditions:

$$n^{\tau}_{ij,r} = n_0 - \frac{(1-\tau_r)cd_{ij,r}}{s^{\tau}_{j,r}} + \theta_{ij,r} \tag{42}$$

$$s^{\tau}_{j,r} = 1 + \alpha \sum_{k=1}^{N_r} n^{\tau}_{jk,r}\, s^{\tau}_{k,r} \tag{43}$$

The total welfare per network is defined as (see above)

$$W^{\tau}_r = \sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\left[\left((n_0+\theta_{ij,r})\, n^{\tau}_{ij,r} - \frac{1}{2}\left(n^{\tau}_{ij,r}\right)^2\right) s^{\tau}_{j,r} - n^{\tau}_{ij,r}(1-\tau_r)cd_{ij,r}\right] \tag{44}$$

As for the social interaction subsidy, we find the subsidy τ^*_r that gives the same aggregate utility W^{τ}_r in network r as the first best W^0_r. 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 equations (42) and (43) and find the subsidy such that $W^{\tau}_r = W^0_r$.

The last three columns in Table 7 display the results. On average, a subsidy level of $\tau = 0.832$ (83.2%) is required for a network to achieve the first best aggregate level of social interactions and social capital.

6.2.3 Comparing the two policies

Finally, it is interesting to compare these two policies at the same given cost. The question is then as follows: Given that the planner has an amount B to spend, which policy should she choose? In order to distribute a total amount of subsidy B to each network, we consider three different schemes. First, we distribute the same amount $B_r = B/R$ for each network (uniform subsidy), where R is the total number of networks ($R = 141$ in our dataset). The second scheme gives an amount proportional to network population N_r. Hence, $B_r = \frac{N_r}{\sum_{r'} N_{r'}} B$. The last subsidy scheme provides an amount proportional to the number of pairs $N_r(N_r-1)$, i.e., $B_r = \frac{N_r(N_r-1)}{\sum_{r'} N'_r(N_{r'}-1)} B$.

We consider two ways of assigning the total amount of subsidy budget. First, we choose the amount of budget that corresponds to the average social interaction subsidy level that achieves the first best level of social interactions:

$$B := B^{\sigma} = \bar{\sigma}^{o}\bar{n}^{o}\sum_{r=1}^{R} N_r(N_r-1), \tag{45}$$

where $\bar{\sigma}^o$ is the average optimal social interaction subsidy level, as obtained in Table 7 (i.e., $\bar{\sigma} = 0.826$) and $\bar{n}^o$ is the average optimal social interaction level, as obtained in Table 3 (i.e., $\bar{n}^o = 3.154$).

Second, we use the amount of budget that corresponds to the average transportation subsidy level to achieve the first best level of social interactions:

$$B := B^{\tau} = \bar{\tau}^{o}c\ \bar{n}^{o}\sum_{r=1}^{R} N_r(N_r-1), \tag{46}$$

where $\bar{\tau}^o$ is the average transportation subsidy rate, i.e., $\bar{\tau}^o = 0.832$ (Table 7).

We proceed as follows. First, we consider the *social-interaction subsidy policy.* We observe $d_{ij,r}$ and N_r in the data and have estimated α, c and n_0. Then, we solve simultaneously equations (39), (40) and (45). We get the different endogenous variables, in particular, the different subsidies σ_r. Then, for each value of σ_r, we calculate the total welfare W_r^{σ} given by (41).

Second, we consider the *transportation subsidy policy.* We observe $d_{ij,r}$ and N_r in the data and have estimated α, c and n_0. Then, we solve simultaneously equations (42), (43) and (46). We get the endogenous variables, in particular, the different subsidies τ_r. Then, for each value of τ_r, we calculate the total welfare W_r^{τ} given by (44).

Our key question is then whether the budget B_σ and B_τ yield higher welfare with the subsidy on travel costs or social interactions. That is, we examine whether $W_r^{\tau} \gtreqless W_r^{\sigma}$. Table 8 shows the results of this analysis by counting the number of networks for which the total welfare is higher under one policy versus the other. In this table, we find that, under the social-interaction subsidy policy, the total welfare is higher for most networks, regardless of the amount of budget we assign (panels A and B) and the type of subsidy scheme (uniform, proportional to N_r and proportional to $N_r(N_r-1)$; rows (1), (2) and (3)).[22] As a result, if a planner has a given amount of money to spend, she should subsidize social interactions and

[22] We also try different values of the total amount to be spent to check whether there are non-linear effects, but the results remain the same regardless of the value of the budget.

not transportation costs because it yields greater improvements of total welfare.[23]

[*Insert Table 8 here*]

7 Concluding remarks

In this paper, we present a behavioral microfoundation for the relationship between geographical distance and social interactions. We characterize the equilibrium in terms of optimal level of social interactions and social capital for a general distribution of individuals in the geographical space. An important prediction of the model is that the 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. Social capital tends to be more concentrated than agents themselves. This result seems to be confirmed by what we observe in real-world cities. Indeed, despite rapid innovation in communication technologies, we still observe an important growth in urbanization (Henderson (2010)), which may highlight the importance of geographical proximity for social exchange (Bailey et al. (2020)). We also show that greater spatial dispersion of agents in the city (which increases trip distances and costs) decreases the incentives to socially interact. As a result, greater spatial dispersion reduces social capital. Because of the externalities that agents exert on each other, we demonstrate that the equilibrium levels of social interactions and social capital are lower than the efficient ones.

When we estimate the model using data on adolescents in the United States we find that, indeed, geographical distance is an hinder to social interactions. Moreover, we determine the exact inefficiencies of the market equilibrium. Interestingly, and surprisingly, we find that there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size. In our empirical context, these inefficiencies are the largest when the network is composed of 8 students. On the contrary, we find that the network that maximizes the average welfare in a network should have 11 students. We then perform two different subsidy policies. Our results suggest that the individuals interact at optimal levels when

[23] Observe that this result is not in contradiction with Proposition 6, which shows that, in order to restore the first-best solutions, social interactions should not be subsidized while transportation costs should be. In Table 8, we are not calculating the subsidy levels that restore the first best. Instead, we are determining which policy leads to a higher total welfare for a given cost. The first best may clearly not be reached. On the contrary, in Table 7, we are calculating the subsidy level that restores the first best for each policy.

either social interactions or transportation costs are subsidized. However, subsidies on social interactions are more effective than subsidies on transportation costs.

Our analysis thus suggests that encouraging social interactions in cities are likely to enhance social welfare, which is a new implication compared to what urban economics usually predicts.[24] In the real-world, there are different ways governments can subsidize social interactions. One natural way is *social mixing* such as the Moving to Opportunity (MTO) programs in the United States where the local government subsidizes housing to allow families to move from poor to richer neighborhoods (see e.g., Katz, Kling, and Liebman (2001), Kling, Liebman, and Katz (2007) and Chetty, Hendren, and Katz (2016)). These programs allow people from different neighborhoods to interact with each other. Other policies that enhance social interactions are those that improve physical environment such as zoning laws and public housing rules (Glaeser and Sacerdote (2000)). For example, Glaeser and Sacerdote (2000) find that individuals in large apartment buildings are more likely to socialize with their neighbors than those living in smaller apartment buildings. Using Facebook data from the United States, Bailey et al. (2018a) document that, at the county level, friendship networks are a mechanism that can propagate house price shocks through the economy via housing price expectations. These types of policies may be particularly important under the view that social interactions promote economic growth because of the nonmarket intellectual spillovers that they generate (Glaeser, 2000; Ioannides, 2013)[25] but also because of the direct effects social interactions have on innovation (Bailey et al. (2018b)) and on the labor market (see e.g., Ioannides and Datcher Loury (2004) or Beaman (2016)).

We believe that more research in this area should be done, especially empirically, in order to be able to better evaluate the exact role of social interactions on the growth and welfare of cities.

[24] In the standard monocentric models (Fujita et al., 1999) and in their multicentric extensions (Fujita and Thisse, 2013), unit travel cost is usually the fundamental parameter that determines the location choices of households within cities, their consumption of housing, land use, and the population size of cities. As a result, transportation policies that reduce commuting costs in the city have been put forward because of their direct impact of these outcomes.

[25] Indeed, as argued by Romer (1986) and Lucas (1988), endogenous economic growth requires increasing returns and without nonmarket intellectual spillovers or some form of externality, increasing returns cannot coexist. The robust relationship between human capital and economic growth has been taken as support for the importance of these intellectual spillovers (Combes and Gobillon (2015)).

References

Ackerberg, Daniel A and Gautam Gowrisankaran. 2006. "Quantifying equilibrium network externalities in the ACH banking industry." *The RAND Journal of Economics* 37 (3):738–761.

Arzaghi, Mohammad and J Vernon Henderson. 2008. "Networking off madison avenue." *The Review of Economic Studies* 75 (4):1011–1038.

Bailey, Michael, Ruiqing Rachel Cao, Theresa Kuchler, and Johannes Stroebel. 2018a. "The economic effects of social networks: Evidence from the housing market." *Journal of Political Economy* 126 (6):2224–2276.

Bailey, Michael, Ruiqing Rachel Cao, Theresa Kuchler, Johannes Stroebel, and Arlene Wong. 2018b. "Social connectedness: Measurement, determinants, and effects." *Journal of Economic Perspectives* 32 (3):259–280.

Bailey, Michael, Patrick Farrell, Theresa Kuchler, and Johannes Stroebel. 2020. "Social connectedness in urban areas." *Journal of Urban Economics* 118:103264.

Barthélemy, Marc. 2011. "Spatial networks." *Physics Reports* 499 (1-3):1–101.

Barwick, Panle Jia, Yanyan Liu, Eleonora Patacchini, and Qi Wu. 2019. "Information, mobile communication, and referral effects." CEPR Discussion Paper No. 13786 .

Bayer, Patrick, Stephen L Ross, and Giorgio Topa. 2008. "Place of work and place of residence: Informal hiring networks and labor market outcomes." *Journal of Political Economy* 116 (6):1150–1196.

Beaman, Lori. 2016. "Social networks and the labor market." In *Oxford Handbook on the Economics of Networks*, edited by Y. Bramoullé, A. Galeotti, and B. Rogers. Oxford: Oxford University Press.

Beckmann, Martin J. 1976. "Spatial equilibrium in the dispersed city." *Environment, regional science and interregional modeling* 127:132–141.

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. "Productive cities: Sorting, selection, and agglomeration." *Journal of Political Economy* 122 (3):507–553.

Belhaj, Mohamed, Sebastian Bervoets, and Frédéric Deroïan. 2016. "Efficient networks in games with local complementarities." *Theoretical Economics* 11 (1):357–380.

Berliant, Marcus, Shin-Kun Peng, and Ping Wang. 2002. "Production externalities and urban configuration." *Journal of Economic Theory* 104 (2):275–303.

Bisztray, M., M. Koren, and A. Szeidl. 2018. "Learning to import from your peers." *Journal of International Economics* 115:242–258.

Bramoullé, Yann, Brian W. Rogers, and Andrea Galeotti. 2016. *The Oxford Handbook of the Economics of Networks*. Oxford University Press.

Brueckner, Jan K and Ann G Largey. 2008. "Social interaction and urban sprawl." *Journal of Urban Economics* 64 (1):18–34.

Büchel, Konstantin and Maximilian von Ehrlich. 2019. "Cities and the structure of social Interactions: Evidence from mobile phone data." Unpublished manuscript, Universität Bern .

Cabrales, Antonio, Antoni Calvó-Armengol, and Yves Zenou. 2011. "Social interactions and spillovers." *Games and Economic Behavior* 72 (2):339–360.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. 2009. "Peer effects and social networks in education." *The Review of Economic Studies* 76 (4):1239–1267.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2016. "The effects of exposure to better neighborhoods on children: New evidence from the Moving to Opportunity experiment." *The American Economic Review* 106 (4):855–902.

Combes, Pierre-Philippe and Laurent Gobillon. 2015. "The empirics of agglomeration economies." In *Handbook of Regional and Urban Economics, Volume 5A*, edited by G. Duranton, V. Henderson, and W. Strange. Amsterdam: Elsevier, 247–348.

Currarini, S., M.O. Jackson, and P. Pin. 2009. "An economic model of friendship: Homophily, minorities, and segregation." *Econometrica* 77 (4):1003–1045.

Duranton, Gilles and Diego Puga. 2015. "Urban Land Use." *Handbook of Regional and Urban Economics* 5:467–560.

Fafchamps, Marcel and Flore Gubert. 2007. "Risk sharing and network formation." *The American Economic Review* 97 (2):75–79.

Fu, Chao and Jesse Gregory. 2019. "Estimation of an Equilibrium Model With Externalities: Post-Disaster Neighborhood Rebuilding." *Econometrica* 87 (2):387–421.

Fujita, Masahisa. 1989. *Urban Economic Theory: Land Use and City Size*. Cambridge University Press.

Fujita, Masahisa, Paul R Krugman, Anthony J Venables, and Massahisa Fujita. 1999. *The Spatial Economy: Cities, Regions and International Trade*. MIT Press.

Fujita, Masahisa and Jacques-François Thisse. 2013. *Economics of Agglomeration: Cities, Industrial Location, and Globalization*. Cambridge University Press.

Glaeser, Edward L. 1999. "Learning in cities." *Journal of urban Economics* 46 (2):254–277.

Glaeser, Edward L. 2000. "The future of urban research: nonmarket interactions." *Brookings-Wharton papers on urban affairs* 1:101–149.

Glaeser, Edward L. and Bruce Sacerdote. 2000. "The social consequences of housing." *Journal of Housing Economics* 9:1–23.

Goldenberg, Jacob and Moshe Levy. 2009. "Distance is not dead: Social interaction and geographical distance in the internet era." *arXiv preprint arXiv:0906.3202* .

Gourieroux, Christian, Alain Monfort, and Eric Renault. 1993. "Indirect inference." *Journal of Applied Econometrics* 8 (S1):S85–S118.

Graham, Bryan S. 2017. "An econometric model of network formation with degree heterogeneity." *Econometrica* 85 (4):1033–1063.

Hellerstein, Judith K, Mark J Kutzbach, and David Neumark. 2014. "Do labor market networks have an important spatial dimension?" *Journal of Urban Economics* 79:39–58.

Hellerstein, Judith K, Melissa McInerney, and David Neumark. 2011. "Neighbors and Coworkers: The Importance of Residential Labor Market Networks." *Journal of Labor Economics* 29 (4):659–695.

Helsley, Robert W and William C Strange. 2007. "Urban interactions and spatial structure." *Journal of Economic Geography* 7 (2):119–138.

———. 2014. "Coagglomeration, clusters, and the scale and composition of cities." *Journal of Political Economy* 122 (5):1064–1093.

Helsley, Robert W and Yves Zenou. 2014. "Social networks and interactions in cities." *Journal of Economic Theory* 150:426–466.

Henderson, J. Vernon. 2010. "Cities and development." *Journal of Regional Science* 50 (1):515–540.

Ioannides, Yannis M. 2013. *From Neighborhoods to Nations: The Economics of Social Interactions*. Princeton: Princeton University Press.

Ioannides, Yannis M. and Linda Datcher Loury. 2004. "Job information networks, neighborhood effects, and inequality." *Journal of Economic Literature* 42 (4):1056–1093.

Jackson, Matthew O, Brian Rogers, and Yves Zenou. 2017. "The Economic Consequences of Social Network Structure." *Journal of Economic Literature* 55 (1):1–47.

Jackson, Matthew O. and Asher Wolinsky. 1996. "A strategic model of social and economic networks." *Journal of Economic Theory* 71 (1):44–74.

Jackson, Matthew O and Yves Zenou. 2015. "Games on networks." In *Handbook of Game Theory, Volume 4*, edited by P. Young and S. Zamir. Amsterdam: Elsevier, 91–157.

Jackson, M.O. 2008. *Social and Economic Networks*. Princeton: Princeton University Press.

Jackson, M.O. and B.W. Rogers. 2005. "The economics of small worlds." *Journal of the European Economic Association* 3:617–627.

Johnson, Cathleen and Robert P Gilles. 2000. "Spatial social networks." *Review of Economic Design* 5 (3):273–299.

Kaltenbrunner, Andreas, Salvatore Scellato, Yana Volkovich, David Laniado, Dave Currie, Erik J Jutemar, and Cecilia Mascolo. 2012. "Far from the eyes, close on the web: impact of geographic distance on online social interactions." In *Proceedings of the 2012 ACM workshop on Workshop on online social networks*. 19–24.

Katz, L.F., J.R. Kling, and J.B. Liebman. 2001. "Moving to opportunity in Boston: Early results of a randomized mobility experiment." *Quarterly Journal of Economics* 116:607—654.

Kling, J.R., J.B. Liebman, and L.F. Katz. 2007. "Experimental analysis of neighborhood effects." *Econometrica* 75 (1):83–119.

König, M., X. Liu, and Y. Zenou. 2019. "R&D networks: Theory, empirics and policy implications." *The Review of Economics and Statistics* 101 (3):476–491.

König, Michael, Claudio Tessone, and Yves Zenou. 2014. "Nestedness in networks: A theoretical model and some applications." *Theoretical Economics* 9:695–752.

Krings, Gautier, Francesco Calabrese, Carlo Ratti, and Vincent D Blondel. 2009. "Urban gravity: a model for inter-city telecommunication flows." *Journal of Statistical Mechanics: Theory and Experiment* 2009 (07):L07003.

Lambiotte, Renaud, Vincent D Blondel, Cristobald De Kerchove, Etienne Huens, Christophe Prieur, Zbigniew Smoreda, and Paul Van Dooren. 2008. "Geographical dispersal of mobile communication networks." *Physica A: Statistical Mechanics and its Applications* 387 (21):5317–5325.

Liben-Nowell, David, Jasmine Novak, Ravi Kumar, Prabhakar Raghavan, and Andrew Tomkins. 2005. "Geographic routing in social networks." *Proceedings of the National Academy of Sciences* 102 (33):11623–11628.

List, J.A., F. Momeni, and Y. Zenou. 2019. "Are estimates of early education programs too pessimistic? Evidence from a large-scale field experiment that causally measures neighbor effects." CEPR Discussion Paper No. 13725 .

Lucas, Robert E. 1988. "On the mechanics of economic development." *Journal of Monetary Economics* 22 (1):3–42.

Lucas, Robert E and Esteban Rossi-Hansberg. 2002. "On the internal structure of cities." *Econometrica* 70 (4):1445–1476.

Manski, C.F. 1993. "Identification of endogenous social effects: The reflection problem." *The Review of Economic Studies* 60 (3):531.

Marmaros, David and Bruce Sacerdote. 2006. “How Do Friendships Form?” *The Quarterly Journal of Economics* 121 (1):79–119.

McPherson, M., L. Smith-Lovin, and J.M. Cook. 2001. “Birds of a feather: Homophily in social networks.” *Annual Review of Sociology* 27 (1):415–444.

Mossay, Pascal and Pierre M. Picard. 2011. “On spatial equilibria in a social interaction model.” *Journal of Economic Theory* 146 (6):2455–2477.

———. 2019. “Spatial segregation and urban structure.” *Journal of Regional Science* 59:480–507.

Ogawa, Hideaki and Masahisa Fujita. 1980. “Equilibrium land use patterns in a nonmonocentric city.” *Journal of regional science* 20 (4):455–475.

Picard, P.M. and Y. Zenou. 2018. “Urban spatial structure, employment and social ties.” *Journal of Urban Economics* 104:77–93.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community.* Simon and Schuster.

Romer, Paul M. 1986. “Increasing Returns and Long-Run Growth.” *The Journal of Political Economy* 94 (5):1002–1037.

Rosenthal, Stuart S. and William C. Strange. 2008. “The attenuation of human capital spillovers.” *Journal of Urban Economics* 46 (2):373—-389.

Sato, Yasuhiro and Yves Zenou. 2015. “How urbanization affect employment and social interactions.” *European Economic Review* 75:131–155.

Schmutte, Ian M. 2015. “Job referral networks and the determination of earnings in local labor markets.” *Journal of Labor Economics* 33 (1):1–32.

Spencer, Barbara J and James A Brander. 1983. “International R&D rivalry and industrial strategy.” *The Review of Economic Studies* 50 (4):707–722.

Zenou, Yves. 2009. *Urban Labor Economics.* Cambridge University Press.

———. 2013. “Spatial versus social mismatch.” *Journal of Urban Economics* 74:113–132.

Appendix

A Proofs

Proof of Proposition 3: We need to show that (*i*) $\text{Disp}(s\lambda) < \text{Disp}(\lambda)$ is equivalent to $\text{Disp}(g\lambda) > \text{Disp}(\lambda)$ and (*ii*) this is true when $x^2\lambda(x)/\int z^2\lambda(z)\mathrm{d}z$ is a mean preserving spread of a symmetric distribution of $\lambda(x)$.

First, note that s_0 is a constant and s and λ are functions of z. One successively gets the following equivalences:

$$\begin{aligned}
&\text{Disp}(s\lambda) < \text{Disp}(\lambda) \\
&\qquad \Leftrightarrow \frac{\int z^2 s\lambda \mathrm{d}z}{\int s\lambda \mathrm{d}z} < \frac{\int z^2 \lambda \mathrm{d}z}{\int \lambda \mathrm{d}z} \\
&\qquad \Leftrightarrow \frac{\int z^2 \left(s_0 - \alpha g\right) \lambda \mathrm{d}z}{\int z^2 \lambda \mathrm{d}z} < \frac{\int \left(s_0 - \alpha g\right) \lambda \mathrm{d}z}{\int \lambda \mathrm{d}z} \\
&\qquad \Leftrightarrow s_0 - \alpha \frac{\int z^2 g\lambda \mathrm{d}z}{\int z^2 \lambda \mathrm{d}z} < s_0 - \alpha \frac{\int g\lambda \mathrm{d}z}{\int \lambda \mathrm{d}z} \\
&\qquad \Leftrightarrow \frac{\int z^2 g\lambda \mathrm{d}z}{\int z^2 \lambda \mathrm{d}z} > \frac{\int g\lambda \mathrm{d}z}{\int \lambda \mathrm{d}z} \\
&\qquad \Leftrightarrow \frac{\int z^2 g\lambda \mathrm{d}z}{\int g\lambda \mathrm{d}z} > \frac{\int z^2 \lambda \mathrm{d}z}{\int \lambda \mathrm{d}z} \\
&\qquad \Leftrightarrow \text{Disp}(g\lambda) > \text{Disp}(\lambda)
\end{aligned}$$

where, for notation convenience, we have dropped the integrals the boundaries $-b$ and b.

Second, by denoting by

$$\mu(z) \equiv \frac{z^2\lambda(z)}{\int_{-b}^{b} w^2\lambda(w)\mathrm{d}w},$$

we can write the last condition $\text{Disp}(g\lambda) > \text{Disp}(\lambda)$ as $\int_{-b}^{b} g\mu\mathrm{d}z - \int_{-b}^{b} g\lambda\mathrm{d}z = \int_{-b}^{b} g\left(\mu - \lambda\right)\mathrm{d}z > 0$. Integrating by part, we obtain the following condition:

$$-\int_{-b}^{b}\left\{\int_{-b}^{z}\left[\mu(x) - \lambda(x)\right]\mathrm{d}x\right\} g'(z)\mathrm{d}z > 0 \tag{A.1}$$

Finally, consider the symmetric spatial distribution $\lambda(x)$ around $x = 0$. Because $\lambda(x)$ is symmetric around $x = 0$, then $g(x) = \int_{-b}^{b} c(x - z)\lambda(z)\mathrm{d}z$ is symmetric around $x = 0$.

Furthermore we know that $g(x)$ is also convex, which implies that $g'(z) > 0$ if and only if $z > 0$. A sufficient condition for inequality (A.1) to be true is that $\int_{-b}^{z}[\mu(x) - \lambda(x)]\,\mathrm{d}x$ is negative for $z > 0$ and positive for $z < 0$. That is, if

$$\int_{-b}^{z} \mu(x)\mathrm{d}x \le \int_{-b}^{z} \lambda(x)\mathrm{d}x, \text{ for } z > 0$$

and the opposite condition for $z < 0$. This condition is satisfied if $\mu(x)$ is a mean preserving spread of the distribution of $\lambda(x)$ around its mean $x = 0$. For example, for a uniform distribution $\lambda(x) = 1/(2b)$, we get

$$\begin{aligned}\int_{-b}^{z} \mu(x)\mathrm{d}x - \int_{-b}^{z} \lambda(x)\mathrm{d}x &= \int_{-b}^{z} \left(\frac{x^2}{\int_{-b}^{b} w^2 \frac{1}{2b}\mathrm{d}w} - 1 \right) \frac{1}{2b}\mathrm{d}x \\ &= -\frac{1}{2} z \left(b^2 - z^2\right)/b^3 < 0\end{aligned}$$

so that $\mu(x)$ is spread of the distribution of $\lambda(x)$. ▮

Proof of Lemma 4: Substituting z for y we can write the Lagrangian function as

$$\begin{aligned}\mathcal{L} = &\int_{-b}^{b}\int_{-b}^{b} \{v[n(x,y)]\,s(y) - n(x,y)c(x-y) + \alpha\chi(x)n(x,y)s(y)\}\,\lambda(y)\lambda(x)\mathrm{d}x\mathrm{d}y \\ &- \int_{-b}^{b} \chi(x)\,[s(x) - 1]\,\lambda(x)\mathrm{d}x\end{aligned}$$

Finally, we note that $\int_{-b}^{b} \chi(x)\,[s(x) - 1]\,\lambda(x)\mathrm{d}x = \int_{-b}^{b} \chi(y)\,[s(y) - 1]\,\lambda(y)\mathrm{d}y$. Substituting the latter expression in the last term in the above expression and multiplying it by $\int_{-b}^{b} \lambda(x)\mathrm{d}x$ $(= 1)$ we get the following Lagrangian function:

$$\mathcal{L} = \int_{-b}^{b}\int_{-b}^{b} \left\{ \begin{array}{c} v\,[n(x,y)]\,s(y) - n(x,y)c(x-y) \\ +\alpha\chi(x)n(x,y)s(y) - \chi(y)\,[s(y) - 1] \end{array} \right\} \lambda(y)\lambda(x)\mathrm{d}x\mathrm{d}y \tag{A.2}$$

We now use variation calculus on the Lagrangian function $\mathcal{L} = \int_{-b}^{b}\int_{-b}^{b} F[n(x,y), s(y), x, y]\lambda(y)\lambda(x)\mathrm{d}x\mathrm{d}y$ where $F(n, s, x, y)$ denotes the integrand in the above curly bracket. It is a differentiable function with partial derivatives F_n' and F_s'. Defining the infinitely small perturbations $\widetilde{n}(x,y)$ and $\widetilde{s}(y)$ on the optimal profiles $n^o(x,y)$ and $s^o(y)$ respectively, we get the variation of the objective function $\mathcal{L}$

$$\begin{aligned}\Delta\mathcal{L}= & \int_{-b}^{b}\int_{-b}^{b} F_n'[n^o(x,y),s^o(y),x,y]\widetilde{n}(x,y)\lambda(y)\lambda(x)\mathrm{d}x\mathrm{d}y \\ & +\int_{-b}^{b}\int_{-b}^{b} F_s'[n^o(x,y),s^o(y),x,y]\widetilde{s}(y)\lambda(y)\lambda(x)\mathrm{d}x\mathrm{d}y\end{aligned}$$

This must be zero for any small perturbations $\widetilde{n}(x,y)$ and $\widetilde{s}(y)$. So, we get $F_n'[n^o(x,y),s^o(y),x,y]=0$ and $\int_{-b}^{b}\left\{F_s'[n^o(x,y),s^o(y),x,y]\right\}\lambda(x)\mathrm{d}x=0$. This gives (16) and (17). ▮

Proof of Proposition 5: Condition (16) yields

$$v'\left[n(x,y)\right]=\frac{c(x-y)}{s(y)}-\alpha\chi(x) \tag{A.3}$$

which gives

$$n^o(x,y)=1-\frac{c(x-y)}{s^o(y)}+\alpha\chi^o(x)$$

under our specification of v. With social capital at y held fixed at the equilibrium level ($s^*(y)=s^o(y)$), this expression is larger than the equilibrium number of visits $n^*(x,y)$ because $\chi^o(x)\geq 0$. The question thus becomes how social capital changes in this efficient allocation.

Inserting (10) in the binding condition (15), we get

$$s^o(x)=1+\alpha\int_{-b}^{b}s^o(z)\lambda(z)\mathrm{d}z-\alpha g(x)+\alpha^2\chi^o(x)\int_{-b}^{b}s^o(z)\lambda(z)\mathrm{d}z$$

We use the same algebraic manipulation leading to expression (8), multiplying both members of the last expression by $\lambda(x)$, integrating them and simplifying to get the value of $\int_{-b}^{b}s^o(x)\lambda(x)\mathrm{d}x$. We then insert this expression in the previous equality and simplify, getting the following closed-form solution for the efficient level of social capital:

$$s^o(x)=1+\alpha\frac{[1+\alpha\chi^o(x)]\left[1-\alpha\int_{-b}^{b}g(z)\lambda(z)\mathrm{d}z\right]}{1-\alpha-\alpha^2\int_{-b}^{b}\chi^o(z)\lambda(z)\mathrm{d}z}-\alpha g(x)$$

If $\chi^o(x)=0$, this yields the equilibrium $s^*(x)$. However, since $\chi^o(x)\geq 0$, the numerator is larger and the denominator is smaller than in the equilibrium. It thus must be that $s^o(x)>s^*(x)$. In turn, this implies that $n^o(x,y)\geq n^*(x,y)$. ▮

B Social capital fixed point

The fixed point in social capital can be computed by rewriting equation (20) as $n_{ij,r}s_{j,r} = (n_0 + \theta_{ij,r})\, s_{j,r} - cd_{ij,r}$ so that (21) becomes

$$s_{j,r} = 1 + \alpha\sum_{k=1}^{N_r} \left[(n_0 + \theta_{jk,r})\, s_{k,r}\right] - \alpha c\sum_{k=1}^{N_r} d_{jk,r}, \tag{B.4}$$

where the last term is the discrete equivalent of the access cost measure $g(x)$ in the model. The system of linear equations (B.4) can be written in vector-matrix form as

$$\mathbf{s}_r = \mathbf{1}_{N_r} + \alpha\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)\mathbf{s}_r - \alpha c\mathbf{D}_r\mathbf{1}_{N_r}, \tag{B.5}$$

where $\mathbf{s}_r$ is the $(N_r \times 1)$ vector with elements $s_{i,r}$, $\mathbf{1}_{N_r}$ is the $(N_r \times 1)$ vector of 1 and $\mathbf{N}_{0,r}$ is an $(N_r \times N_r)$ matrix in which the off-diagonal elements are n_0, and the diagonal elements are zero, while $\mathbf{D}_r$= $(d_{ij,r})$ and $\mathbf{\Theta}_r =$ $(\theta_{ij,r}) = (x_{ij,r}^{\mathrm{T}}\beta + \varepsilon_{ij,r})$ are $(N_r \times N_r)$ matrices. Namely,

$$\mathbf{D}_r = \begin{pmatrix} d_{11,r} & ... & d_{1i,r} & ... & d_{1N_r,r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i1,r} & ... & d_{ii,r} & ... & d_{iN_r,r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{N_r1,r} & ... & d_{N_ri,r} & ... & d_{N_rN_r,r} \end{pmatrix} \text{ and } \mathbf{\Theta}_r = \begin{pmatrix} \theta_{11,r} & ... & \theta_{1i,r} & ... & \theta_{1N_r,r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_{i1,r} & ... & \theta_{ii,r} & ... & \theta_{iN_r,r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_{N_r1,r} & ... & \theta_{N_ri,r} & ... & \theta_{N_rN_r,r} \end{pmatrix}. \tag{B.6}$$

Solving the system of linear equations leads to

$$\mathbf{s}_r^* = \left[\mathbf{I}_{N_r} - \alpha\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)\right]^{-1}\left(\mathbf{I}_{N_r} - \alpha c\mathbf{D}_r\right)\mathbf{1}_{N_r},$$

where $\mathbf{I}_{N_r}$ is the $(N_r \times N_r)$ identity matrix. The matrix $\mathbf{I}_{N_r} - \alpha\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)$ is invertible if $\alpha < \frac{1}{\rho(\mathbf{N}_0 + \mathbf{\Theta}_r)}$, where $\rho\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)$ is the largest eigenvalue of the matrix $\mathbf{N}_0 + \mathbf{\Theta}_r$.

C Robustness checks

To check the robustness of our estimation results, in particular, the estimates of α in the social capital equation, we use an additional auxiliary model. That is, we regress the ith row and jth column element of $\mathbf{N}_r^3$, i.e., $[\mathbf{N}_r^3]_{ij}$, on the regressors in equations (27) and (28) as follows:

$$[\mathbf{N}_r^3]_{ij} = \gamma_{30} + x_{ij,r}^{\mathrm{T}}\gamma_{31} + \gamma_{32}d_{ij,r} + \epsilon_{3,ij,r}. \tag{C.7}$$

We check whether our estimates using two auxiliary equations and three auxiliary equations are close to each other. We use the symmetric specification, used in column (2) of Table 2. The results are shown in column (1) of Table B1. All the parameter estimates, including the social capital parameter α are similar to those shown in column (2) of Table 2.

In addition, we check the robustness of our results to the choice of network size. The structural estimates may be heterogeneous across networks with different sizes. To check this, we restrict the size of networks in the following ways: First, we use networks of size 5–10, and second, we use networks of size 4–9. Columns (2) and (3) of Table B1 show the results. Our results are close to the original estimation results, shown in column (4) of Table 2. In particular, the structural parameters n_0, α and c exhibit virtually the same estimates. Hence, we find that our main estimation results are not sensitive to the choice of network sizes.

Table B1: Robustness checks

	(1)	**(2)**	**(3)**
	Adding $\mathbf{N}^3$	Network size	Network size
	equation	4–9	5–10
Social interaction equation			
β***: social distances***	$\lvert x_i - x_j \rvert$	$\lvert x_i - x_j \rvert$	$\lvert x_i - x_j \rvert$
Female	-0.4448***	-0.4824***	-0.2432***
	(0.0775)	(0.1186)	(0.0835)
Black	-0.2875**	-0.2326**	-0.1974**
	(0.1240)	(0.1089)	(0.0887)
Grade	-0.1558***	-0.1473***	-0.1467***
	(0.0367)	(0.0401)	(0.0279)
Family income	-0.0006***	-0.0010***	-0.0008***
	(0.0002)	(0.0002)	(0.0001)
GPA	-0.1293**	-0.0665	-0.0360
	(0.0560)	(0.0601)	(0.0343)
Parental education	-0.0824***	-0.0455***	-0.0659***
	(0.0132)	(0.0146)	(0.0168)
Two parents	0.0272	-0.0557***	-0.0473**
	(0.0184)	(0.0137)	(0.0233)
Physical development	-0.0277***	-0.0188***	-0.0131*
	(0.0057)	(0.0057)	(0.0071)
Religion practice	-0.0843***	-0.1129***	-0.1640***
	(0.0190)	(0.0268)	(0.0186)
Family size	-0.0158***	-0.0159***	-0.0221***
	(0.0019)	(0.0047)	(0.0033)
Family income refused	-0.5195***	-0.7016***	-0.1988*
	(0.0771)	(0.1001)	(0.1194)
β***: combined levels***	$(x_i + x_j)$	$(x_i + x_j)$	$(x_i + x_j)$
Female	-0.0033***	-0.0040***	-0.0019***
	(0.0003)	(0.0007)	(0.0006)
Black	0.0439***	0.0344***	0.0757***
	(0.0045)	(0.0106)	(0.0212)
Grade	0.0530***	0.0639***	0.0668***
	(0.0117)	(0.0106)	(0.0054)
Family income	0.0001	-0.0006***	0.00002
	(0.0001)	(0.0002)	(0.0002)
GPA	-0.0332***	-0.0149*	-0.0349***
	(0.0086)	(0.0083)	(0.0054)
Parental education	-0.0284***	-0.0173***	-0.0193***
	(0.0072)	(0.0066)	(0.0057)
Two parents	-0.0251**	-0.0388***	-0.0335***
	(0.0110)	(0.0092)	(0.0102)
Physical development	0.0259***	0.0290***	0.0276***
	(0.0093)	(0.0094)	(0.0077)
Religion practice	-0.0011	0.0115***	0.0123***
	(0.0020)	(0.0024)	(0.0047)
Family size	-0.0087***	-0.0068***	-0.0114***
	(0.0016)	(0.0029)	(0.0019)
Family income refused	0.0834***	0.1097***	0.0554*
	(0.0277)	(0.0256)	(0.0308)
n_0	1.9332***	1.5423***	1.3139***
	(0.3257)	(0.2244)	(0.0900)
c (transportation cost)	0.1849***	0.1781***	0.1814***
	(0.0094)	(0.0245)	(0.0069)
Social capital equation			
α	0.0860***	0.0896***	0.0880***
	(0.0149)	(0.0107)	(0.0107)
Number of networks	141	136	79
Number of pupils	753	703	505
Number of directed pairs	1,821	1,596	1,449

Note: We estimate parameters $(n_0, c, \beta^{\mathrm{T}})^{\mathrm{T}}$ in the social interaction equation (20) and subsequent specifications (23)–(22), and parameter α in the social capital equation (21).
Bootstrap standard errors (clustered by networks) in parentheses, *** p<0.01, ** p<0.05, * p<0.1

D Calibration in the policy exercises

Consider equations (39)–(43) in Section 6 and denote them as follows:

$$n_{ij,r} = n_0 + \theta_{ij,r} - \frac{\sigma_r - (1-\tau_r)\, cd_{ij,r}}{s_{j,r}} \tag{D.1}$$

and

$$s_{j,r} = 1 + \alpha \sum_{k=1}^{N_r} n_{jk,r} s_{k,r}$$

where we implement together the two policies. The first equation can be written as

$$n_{ij,r} s_{j,r} = (n_0 + \theta_{ij,r})\, s_{j,r} + \sigma_r - (1-\tau_r)\, cd_{ij,r}$$

so that the second equation becomes

$$s_{j,r} = 1 + \alpha \sum_{k=1}^{N_r} \left[(n_0 + \theta_{jk,r})\, s_{k,r} \right] - \alpha \sum_{k=1}^{N_r} \left[\sigma_r - (1-\tau_r)\, cd_{jk,r} \right] \tag{D.2}$$

where the last term is the discrete equivalent of $g(x)$ in the model. Let us denote the $(N_r \times 1)$ vector $\mathbf{s}_r$ as follows: $\mathbf{s}_r = (s_{1,r}, ..., s_{n,r})^{\mathrm{T}}$. Let us also denote the $(N_r \times N_r)$ matrices as: $\mathbf{D}_r = (d_{ij,r})$ and $\mathbf{\Theta}_r = (\theta_{ij,r})$ as in (B.6). Thus, in vector-matrix form, (D.2) can be written as:

$$\mathbf{s}_r = \mathbf{1}_{N_r} + \alpha\, (\mathbf{N}_0 + \mathbf{\Theta}_r)\, \mathbf{s}_r + \alpha \sigma_r N_r \mathbf{1}_{N_r} - \alpha\, (1-\tau_r)\, c\mathbf{D}_r \mathbf{1}_{N_r}$$

where $\mathbf{1}_{N_r}$ is the $(N_r \times 1)$ vector of 1 and $\mathbf{N}_{0,r}$ is an N by N matrix in which the off-diagonal elements are n_0, and the diagonal elements are zero. Solving this equation leads to:

$$\mathbf{s}_r = \left[\mathbf{I}_{N_r} - \alpha\, (\mathbf{N}_0 + \mathbf{\Theta}_r) \right]^{-1} \left[(1 + \alpha \sigma_r N_r)\, \mathbf{1}_{N_r} - \alpha\, (1-\tau_r)\, c\mathbf{D}_r \mathbf{1}_{N_r} \right]$$

or equivalently

$$\mathbf{s}_r = \left[\mathbf{I}_{N_r} - \alpha\, (\mathbf{N}_0 + \mathbf{\Theta}_r) \right]^{-1} \left[(1 + \alpha \sigma_r N_r)\, \mathbf{I}_{N_r} - \alpha\, (1-\tau_r)\, c\mathbf{D}_r \right] \mathbf{1}_{N_r} \tag{D.3}$$

where $\mathbf{I}_{N_r}$ is the $(N_r \times N_r)$ identity matrix. The matrix $\mathbf{I}_{N_r} - \alpha\, (\mathbf{N}_0 + \mathbf{\Theta}_r)$ is invertible if $\alpha < \frac{1}{\rho(\mathbf{N}_0 + \mathbf{\Theta}_r)}$, where $\rho\, (\mathbf{N}_0 + \mathbf{\Theta}_r)$ is the largest eigenvalue of the matrix $\mathbf{N}_0 + \mathbf{\Theta}_r$. As a result, we could solve the model using (D.1) and (D.3). Observe that $n_{ij,r} > 0$ if $(1 + \theta_{ij,r})\, s_{j,r} > (1 - \tau_r)\, cd_{ij,r}$, $\forall i, j$. A sufficient condition is

$$s_{j,r} > \max_{i} \frac{(1 - \tau_r)\, cd_{ij,r} - \sigma_r}{(1 + \theta_{ij,r})}.$$

Table 1: Data description: individual characteristics

		(1)	(2)	Difference	(3)	Difference	(4)	Difference
Variable	Variable definition	Mean (std.dev)	Mean (std.dev)	[P-value]	Mean (std.dev)	[P-value]	Mean (std.dev)	[P-value]
Female	Dummy variable taking value one if the respondent is female	0.51 (0.50)	0.5 (0.50)	[0.90]	0.51 (0.50)	[0.81]	0.53 (0.50)	[0.77]
Black	Dummy variable taking value one if the respondent is Black or African American. “White” is the reference category	0.23 (0.42)	0.24 (0.43)	[0.10]	0.20 (0.40)	[0.25]	0.18 (0.38)	[0.29]
Student grade	Grade of student in the current year, range 7 to 12	9.67 (1.63)	9.49 (1.62)	[0.77]	9.47 (1.61)	[0.58]	9.25 (1.67)	[0.26]
Grade Point Average	Grades defined from “A”=4 to “D and lower”=0. Average of grades in English, math, science and history is taken	2.75 (0.77)	2.78 (0.76)	[0.29]	2.83 (0.75)	[0.12]	2.86 (0.75)	[0.28]
Religion practice	Answer to the question “In the past 12 months, how often did you attend religious services?”. Coded as 1=“once a week or more”, 2=“once a month or more, but less than once a week”, 3=“once a month”, 4=“never”	2.44 (1.44)	2.38 (1.41)	[0.47]	2.38 (1.41)	[0.28]	2.35 (1.38)	[0.08]
Physical development	Answer to the question “How advanced is your physical development compared to other boys your age?”. Coded as 1=“I look younger than most”, 2=“I look younger than some”, 3=“I look average”, 4=“I look older than some”, 5=“I look older than most”	3.19 (1.13)	3.23 (1.12)	[0.18]	3.29 (1.10)	[0.54]	3.36 (1.12)	[0.58]
Two parents	Dummy variable taking value one if the respondent lives in a household with two parents (both biological and non biological) that are married Two parent	0.66 (0.47)	0.68 (0.47)	[0.93]	0.71 (0.45)	[0.55]	0.74 (0.44)	[0.21]
Parental education	Schooling level of the (biological or non-biological) parent who is living with the child, coded as 1 =“never went to school,” 2 =“some school” and “less than high school”, 3= “high school graduate”, “GED”, “went to a business, trade or vocational school”, “some college”, 4 = “graduated from college or a university”, 5 = “professional training beyond a four-year college” If both parents are in the household, the maximum level of schooling is considered	3.09 (0.97)	3.11 (0.95)	[0.45]	3.19 (0.92)	[0.31]	3.15 (0.98)	[0.90]
Family income	Family income in thousands of dollars	40.72 (50.76)	39.93 (50.32)	[0.16]	43.36 (56.77)	[0.64]	48.41 (70.46)	[0.68]
Family size	Number of people living in the household	3.61 (1.67)	3.52 (1.51)	[0.80]	3.42 (1.39)	[0.51]	3.44 (1.30)	[0.77]
Family income missing	Dummy variable taking value one 1 if family income of the respondent is missing	0.91 (0.29)	0.11 (0.31)	[0.76]	0.10 (0.30)	[0.45]	0.10 (0.31)	[0.33]
N.obs		20,745	12,761		4,449		753	

Note: (1): original sample, (2): sample with geocoded information, (3): Sample with social-interaction information, (4) Sample in networks of size 4–10. T-tests for differences in means are performed. P-values are reported squared brackets. Differences are computed with respect to the larger sample in the previous column.

Table 2: Structural estimation results

	(1) Directed	**(2) Undirected**
Social interaction equation		
β***: social distances***	$(x_i - x_j)$	$\lvert x_i - x_j \rvert$
Female	-0.0091	-0.4424*
	(0.1436)	(0.2360)
Black	-0.0912	-0.3492**
	(0.1133)	(0.1708)
Grade	-0.0120	-0.1147*
	(0.0429)	(0.0653)
Family income	0.0001	-0.0006*
	(0.0002)	(0.0003)
GPA	0.0144	-0.1413**
	(0.0443)	(0.0690)
Parental education	0.0040	-0.0502
	(0.0211)	(0.0375)
Two parents	-0.0494	-0.0672*
	(0.0360)	(0.0361)
Physical development	-0.0175**	-0.0184*
	(0.0071)	(0.0100)
Religion practice	-0.0157	-0.0721*
	(0.0231)	(0.0394)
Family size	-0.0103**	-0.0151*
	(0.0057)	(0.0080)
Family income refused	0.0136	-0.4096*
	(0.1155)	(0.2150)
β***: combined levels***	$(x_i + x_j)$	$(x_i + x_j)$
Female	-0.0027***	-0.0026*
	(0.0010)	(0.0014)
Black	0.0320***	0.0325*
	(0.0136)	(0.0183)
Grade	0.0592**	0.0722**
	(0.0242)	(0.0331)
Family income	0.0003	0.0003*
	(0.0004)	(0.00018)
GPA	-0.0322***	-0.0292**
	(0.0119)	(0.0147)
Parental education	-0.0227**	-0.0292*
	(0.0099)	(0.0118)
Two parents	-0.0338***	-0.0322*
	(0.0140)	(0.0185)
Physical development	0.0211**	0.0230
	(0.0106)	(0.0166)
Religion practice	0.0084**	0.0081*
	(0.0035)	(0.0043)
Family size	-0.0081***	-0.0088*
	(0.0027)	(0.0046)
Family income refused	0.0848**	0.0876*
	(0.0359)	(0.0512)
n_0	1.0115***	1.3981*
	(0.3906)	(0.6774)
c (transportation cost)	0.2185***	0.1866**
	(0.0279)	(0.0904)
Social capital equation		
α	0.0810***	0.0851**
	(0.0218)	(0.0400)
Number of networks	141	141
Number of pupils	753	753
Number of directed pairs	1,821	1,821

Note: We estimate parameters $(n_0, c, \beta^{\mathrm{T}})^{\mathrm{T}}$ in the social interaction equation (20) and subsequent specifications (23)–(22), and parameter α in the social capital equation (21).
Bootstrap standard errors (clustered by networks) in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 3: Social interactions and social capital: Optimal level vs. observed level

Social interactions					Social capital				
Optimal level	Observed level	Average difference	Minimum difference	Maximum difference	Optimal level	Observed level	Average difference	Minimum difference	Maximum difference
3.154	1.262	1.892	0.532	3.757	2.448	1.614	0.834	-2.354	11.668

Note: The statistics are computed using the network-level average social interactions and social capital from 141 networks. For example, the largest difference between the average levels of optimal and observed social interactions is 3.757. Note that these statistics differ from pair-level averages.

The observed level of social capital is augmented using equation (24).

We report the average results over 100 simulations.

Table 4: Difference between optimal level and observed level of social interactions

	Optimal−Observed (social interactions)				
	(1)	(2)	(3)	(4)	(5)
Network population	1.268***	1.278***	1.370***	1.121	1.502**
	(0.166)	(0.168)	(0.173)	(0.788)	(0.682)
Network population^2	-0.084***	-0.085***	-0.089***	-0.075**	-0.089***
	(0.013)	(0.013)	(0.013)	(0.038)	(0.033)
Avg. geographic distance		0.0039	0.0009	0.0012	-0.0025
		(0.0075)	(0.0074)	(0.0074)	(0.0069)
Avg. degree centrality			-0.591***	-0.653*	-0.815**
			(0.179)	(0.392)	(0.356)
Std.dev. of degree centrality			0.168	-0.130	-0.111
			(0.160)	(0.515)	(0.462)
Avg. eigenvector centrality				-4.128	0.938
				(7.993)	(7.000)
Clustering coefficient				0.088	0.501
				(0.975)	(0.916)
Diameter				-0.106	-0.062
				(0.178)	(0.159)
Female fraction					0.103
					(0.154)
Black fraction					-0.161
					(0.111)
Avg. student grade					-0.062***
					(0.022)
Avg. GPA					0.221***
					(0.079)
Avg. level of physical development					-0.049
					(0.061)
Avg. level of religion practice					-0.107**
					(0.045)
Avg. family size					0.085
					(0.0515)
Fraction of students with two parents					-0.040
					(0.163)
Avg. level of parent education					0.074
					(0.072)
Avg. family income					-0.0030**
					(0.0013)
Fraction family income refused					-1.012***
					(0.232)
Constant	-2.271***	-2.324***	-1.808***	1.555	-2.017
	(0.492)	(0.504)	(0.514)	(6.286)	(5.517)
Observations	141	141	141	141	141
R-squared	0.462	0.463	0.503	0.506	0.673

Note: The outcome variable is the average difference between optimal level and observed level of social interactions ($N^{opt} - N^{obs}$) over 100 simulations for each network.
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 5: Difference between optimal level and observed level of social capital

	Optimal−Observed (social capital)				
	(1)	(2)	(3)	(4)	(5)
Network population	-4.545***	-4.523***	-4.410***	1.239	1.163
	(0.641)	(0.647)	(0.691)	(3.119)	(3.099)
Network population^2	0.290***	0.288***	0.280***	0.017	0.009
	(0.050)	(0.051)	(0.053)	(0.149)	(0.148)
Avg. geographic distance		0.009	0.010	0.010	0.023
		(0.029)	(0.029)	(0.029)	(0.032)
Avg. degree centrality			0.352	-1.010	-1.262
			(0.717)	(1.553)	(1.615)
Std.dev. of degree centrality			-0.618	-0.153	-0.588
			(0.641)	(2.040)	(2.100)
Avg. eigenvector centrality				58.36*	52.08
				(31.63)	(31.80)
Clustering coefficient				4.000	5.312
				(3.860)	(4.160)
Diameter				0.039	-0.109
				(0.704)	(0.723)
Female fraction					0.222
					(0.700)
Black fraction					0.266
					(0.503)
Avg. student grade					-0.176*
					(0.098)
Avg. GPA					-0.253
					(0.359)
Avg. level of physical development					0.180
					(0.277)
Avg. level of religion practice					0.317
					(0.206)
Avg. family size					0.016
					(0.234)
Fraction of students with two parents					0.129
					(0.739)
Avg. level of parent education					-0.317
					(0.328)
Avg. family income					0.001
					(0.006)
Fraction family income refused					2.565**
					(1.054)
Constant	16.07***	15.95***	15.50***	-29.62	-23.76
	(1.899)	(1.946)	(2.055)	(24.88)	(25.06)
Observations	141	141	141	141	141
R-squared	0.474	0.475	0.478	0.493	0.558

Note: The outcome variable is the difference between optimal level and observed level of social capital ($S^{opt} - S^{obs}$) over 100 simulations for each network.
The observed level of social capital is augmented using equation (24).
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 6: Optimal network design: average welfare and number of students

	(1) Welfare	(2) Welfare	(3) Welfare	(4) Welfare	(5) Welfare
Network population	-1.773	-3.310	-0.446	18.61	20.87**
	(2.793)	(2.567)	(2.539)	(11.32)	(10.45)
Network population^2	0.010	0.136	-0.0150	-0.817	-0.939*
	(0.218)	(0.200)	(0.193)	(0.542)	(0.498)
Avg. geographic distance		-0.615***	-0.679***	-0.670***	-0.667***
		(0.115)	(0.108)	(0.107)	(0.106)
Avg. degree centrality			-12.49***	-19.61***	-21.72***
			(2.637)	(5.634)	(5.445)
Std.dev. of degree centrality			0.751	-10.03	-12.91*
			(2.356)	(7.400)	(7.082)
Avg. eigenvector centrality				134.3	138.5
				(114.8)	(107.2)
Clustering coefficient				17.57	24.10*
				(14.01)	(14.03)
Diameter				-4.428*	-4.931**
				(2.555)	(2.436)
Controls	No	No	No	No	Yes
Observations	141	141	141	141	141
R-squared	0.112	0.264	0.373	0.406	0.553

Note: The outcome variable is the simulated average welfare (AW_r), averaged over 100 simulations for each network.
Control variables include the averages of the social distances and the combined levels used in structural estimation. See Tables 4–5.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Policy levels for optimal outcomes

(1) Subsidizing social interactions: σ			(2) Subsidizing transportation costs: τ		
Average	Minimum	Maximum	Average	Minimum	Maximum
0.826	0.000	22.003	0.832	0.318	1.000

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 in (31)–(33).
We report the average results over 100 simulations.

Table 8: Comparison of two policies

Panel A: Budget corresponding to the average (optimal) social interaction subsidy level

	Number of networks that lead to higher welfare for each policy	
Subsidy schemes	Policy: σ	Policy: τ
(1) Uniform subsidy amount for each network	119	22
(2) Subsidy proportional to N_r	136	5
(3) Subsidy proportional to $N_r(N_r-1)$	119	22

Panel B: Budget corresponding to the average (optimal) transportation subsidy level

	Number of networks that lead to higher welfare for each policy	
Subsidy schemes	Policy: σ	Policy: τ
(1) Uniform subsidy amount for each network	119	22
(2) Subsidy proportional to N_r	136	5
(3) Subsidy proportional to $N_r(N_r-1)$	120	21

Note: 141 networks.
The median number of networks over 100 simulations, which lead to higher welfare for each policy is reported.