Urban Interactions*

Jun Sung Kim† Eleonora Patacchini‡ Pierre M. Picard§
Yves Zenou¶

November 2, 2017

Abstract

This paper studies social-tie formation when individuals care about the geographical location of other individuals. In our model, the intensity of social interactions can be chosen at the same time as friends. We characterize the equilibrium in terms of both social interactions and social capital (the value of social interactions offered by each agent) for a general distribution of individuals in the urban geographical space. We show that greater geographical dispersion decreases the incentives to socially interact. 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 less than the social first best optimum. Our policy analysis suggests that, given the same cost, subsidizing social interactions yields a higher total welfare than subsidizing transportation costs. (JEL codes: R1, R23, Z13)

*We are grateful to Jan K. Brueckner and Paco Maruhenda for insighful discussions and inspiring comments. Yves Zenou gratefully acknowledges the financial support from the French National Research Agency grant ANR-12-INEG-0002.

†Monash University, Australia. E-mail: jun.sung.kim@monash.edu.

‡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 IFN, 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 to the construction 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 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 (implicitly) 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 distribution function of a general form. Each agent decides on 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 structurally estimate the model using data from the National Longitudinal Survey of Adolescent Health (Add Health) in the United States. While the Add Health data has been used extensively for its social network information based on friends' nominations

[1] For recent overviews, see Ioannides (2013), Jackson and Zenou (2015), and Jackson et al. (2017).

(see, e.g. Calvó-Armengol et al. (2009), Currarini et al. (2010), Lin (2010), Bifulco et al. (2011), Card and Giuliano (2013)), this dataset contains other unique information that has not been exploited before. Indeed, Add Health also provides the longitude and latitude coordinates of the residential location of each respondent. As a result, it is possible to obtain information not only on the precise geometry of social contacts and the strength of their social interactions, but also on the geographical distance between them. The challenges in our empirical analysis are that the intensity of social interactions can be chosen at the same time as friends, and that the interaction value offered from a friend (social capital) is unobserved to the econometrician. We address these challenges using the method of simulated moments (MSM) to predict the social capital. This method is widely employed in the empirical industrial organization literature to estimate micro behavioral models where the choices of the agents for each value of the parameter vector are unobserved (see, e.g. Berry (1992)). Very recently it has also been used in the network literature to deal with the complexity of the decision environment when network formation is endogenous (see Banerjee et al. (2013), Ciliberto and Tamer (2009), Sheng (2015), and Badev (2017)). To the best of our knowledge, it has never been used to estimate urban models in presence of an unobserved endogenous variable.[2]

The estimation results provide evidence supporting our theory. We find that transportation costs (and hence geographic distance), social distances, and combined levels of socio-demographic characteristics are all important factors in determining the intensity of social interactions. Students interact with one another far less than the socially optimal level, and thus accumulate less social capital. We show that these inefficiencicies can be explained by the geographic dispersion of students but also by the size of the network. In fact, there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the size of the network so that there are less inefficiencies in both small or large networks compared to networks of average size (around 8 students in our data). By subsidizing social interactions or transportation costs, policymakers can then improve the intensity of social interactions. We find that, at a given cost, subsidizing social interactions is more effective than subsidizing transportation costs in the sense that it leads to a higher welfare.

Our theoretical framework provides a bridge between two literatures: the traditional ur-

[2] The only (recent) paper using MSM in this field is Geyer (2017), which accounts for the endogeneity of housing prices and selection of households into communities.

ban 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 et al., 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 et al. (2002) show the emergence of a unique centre 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 to explain 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. On the other hand, 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); Mossay et al. (2013), Helsley and Zenou (2014), Sato and Zenou (2015)) 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 on the relevance of geographical location for social interactions in *real world* networks are almost nonexistent (see Ioannides

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

[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 characterize all the possible equilibria. See Jackson (2008) for a discussion of these issues.

(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 et al. (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] Another strand of related literature uses geographic proximity as a proxy for social interactions. Most notably, Bayer et al. (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.[7] To the best of our knowledge there is no study using data on the precise geometry of individual social contacts and the geographical distance between them.

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 model can be found in Appendix A.

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: $\int_{-b}^{b} \lambda(y)\mathrm{d}y = 1$.

[6] See also Fafchamps and Gubert (2007) who show that geographic proximity is a strong correlate of risk-sharing networks and Rosenthal and Strange (2008), Arzaghi and Henderson (2008), Bisztray et al. (2017) and List et al. (2017) who find that knowledge and productivity spillovers are important but decay sharply with distance.

[7] Hellerstein et al. (2011, 2014) and Schmutte (2015) build on the same assumption using matched employer-employee data with residential information.

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

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

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

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

This expression assumes decreasing returns to the frequency of interactions with a given agent; it even assumes negative returns (saturation) above $n = 1$.

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

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

The first constant term (normalized to 1) represents the idiosyncratic interaction value that 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 (2) 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

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

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

[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.

pointwise differentiation yields:

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

Using (1), this is equivalent to:

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

So, the optimal number of interactions is equal to:

$$n^*(x,y) = 1 - \frac{c(x-y)}{s(y)} \tag{3}$$

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.

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

A sufficient condition for this inequality to hold is

$$\min_y s(y) > c(2b) \tag{4}$$

Let us define the *access cost measure* as

$$g(y) \equiv \int_{-b}^{b} c(y-z)\lambda(z)\mathrm{d}z, \tag{5}$$

which is lower than the maximum travel cost $c(2b)$. By plugging (3) into (2) and using (5), we obtain the equilibrium level of social capital $s^*(y)$, which is given by:

$$s^*(y) = 1 + \alpha\int_{-b}^{b} s(z)\lambda(z)\mathrm{d}z - \alpha g(y). \tag{6}$$

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:

$$\int_{-b}^{b} s(z)\lambda(z)\mathrm{d}z = \frac{1}{1-\alpha}\left[1-\alpha\int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z\right]. \tag{7}$$

Inserting this result into (6) yields a closed-form solution for the equilibrium social capital given by:

$$s^*(y) = s_0 - \alpha g(y), \tag{8}$$

where

$$s_0 = \frac{1-\alpha^2\int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z}{1-\alpha}, \tag{9}$$

and where $g(y)$ is defined by (5). 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 (4), a sufficient condition is

$$s_0 - \alpha\left[\max_y g(y)\right] > c(2b) \tag{10}$$

To summarize,

Proposition 1 *Assume $0 < \alpha < 1$ and (10). Then, there exists a unique equilibrium $(n^*(x,y), s^*(y))$, defined for all x, y, such that*

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

and

$$s^*(y) = \frac{1-\alpha^2\int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z}{1-\alpha} - \alpha\int_{-b}^{b} c(y-z)\lambda(z)\mathrm{d}z \tag{11}$$

Let us discuss the properties of the equilibrium social capital $s^*(y)$, defined in (11),[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

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

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 $s(y)$ 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 $n^*_\alpha(y,z)$ denotes 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 (6). Differentiating the latter expression with respect to α leads to:

$$s_\alpha(y) = \int_{-b}^{b} s(z)\lambda(z)\mathrm{d}z - g(y) + \alpha \int_{-b}^{b} s_\alpha(z)\lambda(z)\mathrm{d}z. \tag{12}$$

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 (7) into (12) 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 (8), 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"

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

over its mean value, i.e.

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

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 $\text{Disp}(s\lambda) < \text{Disp}(\lambda)$. Using (8), it is shown in the proof of Proposition 3 in Appendix A that this is equivalent to $\text{Disp}(g\lambda) > \text{Disp}(\lambda)$. 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;*

(*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.

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 (10), 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.

2.4 Linear travel costs and uniform distribution of agents

Assume now that there is a uniform distribution of agents in the city so that $\lambda(x) = 1/2b$, which implies that $\int_{-b}^{b} \lambda(y)\mathrm{d}y = 1$. Assume also as above that travel costs are linear so that

$c(x) = c_1 |x|$ where $c_1 > 0$. It is then straightforward to show (see Appendix A) that:

$$n^*(x,y) = 1 - \frac{c|x-y|}{s^*(y)} \tag{13}$$

and

$$s^*(y) = 1 + \frac{\alpha}{2b}\int_{-b}^{b} n^*(y,z)s^*(z)\mathrm{d}z \tag{14}$$

Compared to the general case (Proposition 1), when agents are uniformly distributed in the city and travel costs are linear, we see more clearly how the size of the city, $2b$, affects the social capital $s^*(y)$ and the frequency of interactions $n^*(x,y)$.

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

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

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

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

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 (1). 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. When the city is uniform, the utility function of individual x can be written as:

$$\begin{aligned} U(x) = S(x) - C(x) &= \int_{-b}^{b} \left\{v\left(n(x,y)\right) s(y) - n(x,y)c(x-y)\right\} \lambda(y)\mathrm{d}y \\ &= \frac{1}{2b}\int_{-b}^{b} \left[\left(\left(n_0 + \theta(x,y)\right) n\left(x,y\right) - \frac{1}{2}\left[n\left(x,y\right)\right]^2\right) s(y) - n(x,y)c\left|x-y\right|\right] \mathrm{d}y. \end{aligned} \tag{18}$$

By pointwise differentiating $U(x)$ 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 (14), which is equal to:

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

Let us assume we observe data from R networks ($r = 1, ..., R$), each comprised of 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:

$$n^*_{ij,r} = n_0 - \frac{cd_{ij,r}}{s^*_{j,r}} + \theta_{ij,r}, \tag{19}$$

and

$$s^*_{j,r} = 1 + \frac{\alpha}{2b_r}\sum_{k=1}^{N_r} n^*_{jk,r}s^*_{k,r}, \tag{20}$$

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

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. We discuss this in Appendix B.

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

$$\theta_{ij,r} = \sum_{m=1}^{M} \beta_m |x_{i,m,r} - x_{j,m,r}| + \sum_{m=1}^{M} \beta_{M+m}(x_{i,m,r} + x_{j,m,r}) + \varepsilon_{ij,r}, \tag{21}$$

where negative values in the vector $(\beta_1, \cdots, \beta_M)^T$ capture *homophily* effects (associated with smaller socio-economic distance $|x_{i,m,r} - x_{j,m,r}|$), and $(\beta_{M+1}, \cdots, \beta_{2M})^T$ 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 et al., 2001; Currarini et al., 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:

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

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

Technical note to ease the estimations of the fixed point Consider (19) and (20). The first equation (19) can be written as:

$$n_{ij,r} s_{j,r} = (n_0 + \theta_{ij,r})\, s_{j,r} - c d_{ij,r}, \tag{23}$$

so that (20) becomes

$$s_{j,r} = 1 + \frac{\alpha}{2b_r} \sum_{k=1}^{N_r} [(n_0 + \theta_{jk,r})\, s_{k,r}] - \frac{\alpha c}{2b_r} \sum_{k=1}^{N_r} d_{jk,r}, \tag{24}$$

where the last term is the discrete equivalent of $g(x)$ in the model. Let us solve the fixed point in (20). Let us denote the $(N_r \times 1)$ vector $\mathbf{s}_r$ by: $\mathbf{s}_r = (s_{1,r}, ..., s_{n,r})^{\mathrm{T}}$. Let us also denote

the $(N_r \times N_r)$ matrices by: $\mathbf{\Delta}_r = (d_{ij,r})$ and $\mathbf{\Theta}_r = (\theta_{ij,r}) = (x_{ij,r}^T \beta + \varepsilon_{ij,r})$. In other words,

$$\mathbf{\Delta}_r = \begin{pmatrix} d_{11,r} & \dots & d_{1i,r} & \dots & d_{1N_r,r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i1,r} & \dots & d_{ii,r} & \dots & d_{iN_r,r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{N_r1,r} & \dots & d_{N_ri,r} & \dots & d_{N_rN_r,r} \end{pmatrix} \text{ and } \mathbf{\Theta}_r = \begin{pmatrix} \theta_{11,r} & \dots & \theta_{1i,r} & \dots & \theta_{1N_r,r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_{i1,r} & \dots & \theta_{ii,r} & \dots & \theta_{iN_r,r} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \theta_{N_r1,r} & \dots & \theta_{N_ri,r} & \dots & \theta_{N_rN_r,r} \end{pmatrix}. \tag{25}$$

Thus, in vector-matrix form, (24) can be written as:

$$\mathbf{s}_r = \mathbf{1}_{N_r} + \frac{\alpha}{2b_r} \left(\mathbf{N}_0 + \mathbf{\Theta}_r \right) \mathbf{s}_r - \frac{\alpha c}{2b_r} \mathbf{\Delta}_r \mathbf{1}, \tag{26}$$

where $\mathbf{1}_{N_r}$ is the $(N_r \times 1)$ vector of 1 and $\mathbf{N}_0$ 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

$$\mathbf{s}_r^* = \left[\mathbf{I}_{N_r} - \frac{\alpha}{2b_r} \left(\mathbf{N}_0 + \mathbf{\Theta}_r \right) \right]^{-1} \left(\mathbf{I}_{N_r} - \frac{\alpha c}{2b_r} \mathbf{\Delta}_r \right) \mathbf{1}_{N_r}, \tag{27}$$

where $\mathbf{I}_{N_r}$ is the $(N_r \times N_r)$ identity matrix. The matrix $\mathbf{I}_{N_r} - \frac{\alpha}{2b_r} \left(\mathbf{N}_0 + \mathbf{\Theta}_r \right)$ is invertible if $\frac{\alpha}{2b_r} < \frac{1}{\rho(\mathbf{N}_0 + \mathbf{\Theta}_r)}$, where $\rho \left(\mathbf{N}_0 + \mathbf{\Theta}_r \right)$ is the largest eigenvalue of the matrix $\mathbf{N}_0 + \mathbf{\Theta}_r$.

Estimation Besides agents' characteristics $x_{ij,r}$, we assume that the data provide:

- $n_{ij,r}^*$, the intensity of social interactions between agents i and j in network r
- $d_{ij,r}$, the geographical distance between agents i and j in network r.
- $2b_r$, the maximum geographical distance between two agents in network r, i.e. $2b_r = \max d_{ij,r}$.

Using this information, we need to recover α, β, c, n_0, and the equilibrium social capital, $s_{j,r}^*$. We employ the method of simulated moments (MSM) proposed by McFadden (1989) and Pakes and Pollard (1989).

The objective of MSM estimation is to find the parameter vector that provides the simulated level of social interactions that best matches the observed level of social interactions. In addition, we match the eigenvector centrality (in lieu of social capital) based on the simulated social interactions matrix $\hat{\mathbf{N}}_r^* = (\hat{n}_{ij,r}^*)$ with the eigenvector centrailty based on the

observed social interactions matrix $\mathbf{N}_r^* = (n_{ij,r}^*)$. The eigenvector centrality is recursively defined as is social capital, and hence matching them helps identification of the parameter in the social capital equation. We explain this with more detail below.

The estimation procedure is as follows. First, recall the equations (23) and (27). We define

$$\mathbf{s}^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \theta) \equiv \left[\mathbf{I}_{N_r} - \frac{\alpha}{2b_r}\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)\right]^{-1} \left(\mathbf{I}_{N_r} - \frac{\alpha c}{2b_r}\mathbf{\Delta}_r\right)\mathbf{1}_{N_r}, \tag{28}$$

where $\mathbf{X}_r$ and $\mathcal{E}_r$, which make up $\mathbf{\Theta}_r$, are the matrices of observed and unobserved pair-level socio-economic characteristics for network r, defined similarly to $\mathbf{\Delta}_r$ in (25).[12]

We use θ to denote the vector of all parameters, i.e. $\theta = (n_0, \alpha, c, \beta^{\mathrm{T}}, \sigma_\varepsilon^2)^{\mathrm{T}}$, where σ_ε^2 is the variance of ε. Let $s_j^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \theta)$ be the jth element of $\mathbf{s}^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \theta)$, i.e. social capital of j. By plugging $s_j^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \theta)$ into (23), $n_{ij,r}^*$ can be expressed as follows:

$$n_{ij,r}^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \theta) = n_0 - \frac{cd_{ij,r}}{s_j^*(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r; \theta)} + x_{ij}^{\mathrm{T}}\beta + \varepsilon_{ij}. \tag{29}$$

Now, we draw T sets of simulation errors $\varepsilon_{ij,r}^{(t)}$, $t = 1, \cdots, T$ for all pairs and all networks. These sets of errors will be fixed for the entire estimation process. Next, we compute social capital $\mathbf{s}^{(t)}$ and predict the intensity of social interactions $\hat{n}_{ij,r}^{(t)}$ for each set of errors using equation (28) and (29). Then, the prediction error is given by

$$\begin{aligned}
\hat{\nu}_{ij,r} &= n_{ij,r}^* - \frac{1}{T}\sum_{t=1}^{T} \hat{n}_{ij,r}^{(t)} \\
&= n_{ij,r}^* - \frac{1}{T}\sum_{t=1}^{T} \Big(n_0 - \frac{cd_{ij,r}}{\hat{s}_j(\mathbf{X}_r, \mathbf{\Delta}_r, b_r, \mathcal{E}_r^{(t)}; \theta)} + x_{ij}^{\mathrm{T}}\beta + \varepsilon_{ij,r}^{(t)}\Big),
\end{aligned} \tag{30}$$

where $\mathcal{E}_r^{(t)}$ is the matrix of the tth set of simulation errors. We use $\hat{s}$ instead of s^* since the former is associated with simulation errors. The prediction error $\hat{\nu}_{ij,r}$ is uncorrelated with exogenous data $x_{ij,r}$ and $d_{ij,r}$ at the true parameter value θ_0 (e.g. Berry, 1992). That is,

$$E(\hat{\nu}_{ij,r} | \mathbf{X}_r, \mathbf{\Delta}_r; \theta = \theta_0) = 0$$

[12] There are a total of $2M$ (i.e. M for the social distances and another M for the combined levels) number of N_r by N_r matrices of $\mathbf{X}_{m,r}$, $m = 1, \cdots, 2M$. Hence, $\mathbf{\Theta}_r = \sum_{m=1}^{2M} \beta_m \mathbf{X}_{m,r} + \mathcal{E}_r$. With a slight abuse of notations, we use $\mathbf{X}_r$ to collect $\mathbf{X}_{m,r}$, $m = 1, \cdots, 2M$.

From this, we have

$$E(\hat{\nu}_{ij,r}) = 0,\ E(\hat{\nu}_{ij,r}x_{ij,r}) = 0 \text{ and } E(\hat{\nu}_{ij,r}d_{ij,r}) = 0. \tag{31}$$

From this, we can construct $(2M + 2)$ moment conditions. However, we have $(2M + 4)$ parameters to estimate, so the model is still under-identified.

To ensure identification, we utilize the relation between social capital and the eigenvector centrality of social interactions matrix $\mathbf{N}_r^*$. Recall the social capital equation (20)

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

and compare it with the eigenvector centrality $EC_{j,r}$ defined as (see, e.g. Jackson (2008))

$$EC_{j,r} = \frac{1}{\lambda}\sum_{k=1}^{N_r} n_{jk,r}^* EC_{k,r}, \tag{32}$$

where λ is the largest eigenvalue of $\mathbf{N}_r^*$. The eigenvector centrality is a reasonable proxy for social capital because it is recursively defined, as social capital, with respect to the social interactions matrix $\mathbf{N}_r^*$. Therefore, if the model precisely predicts the data, the eigenvector centrality from $\mathbf{N}_r^*$ and the eigenvector centrality from the predicted social interaction matrix, say $\hat{\mathbf{N}}_r^*$, must be close to each other. Moreover, matching the eigenvector centralities helps identification of the parameter α which appears only in the social capital equation (20). We define another $N_r \times 1$ vector of predicted errors $\hat{\xi}_r$ such that

$$\hat{\xi}_r = EC_r - \frac{1}{T}\sum_t \widehat{EC}_r^{(t)}, \tag{33}$$

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

These prediction errors are mean independent of $x_{ij,r}$ and $d_{ij,r}$ at the true parameter value θ_0. That is,

$$E(\hat{\xi}_r|\mathbf{X}_r, \mathbf{\Delta}_r; \theta = \theta_0) = 0$$

From the independence between $\hat{\xi}$ and the observed variables $x_{ij,r}$ and $d_{ij,r}$, we have $(2M+2)$

additional moment conditions.

$$E(\hat{\xi}_r) = 0,\ E(\hat{\xi}_r x_{ij,r}) = 0 \text{ and } E(\hat{\xi}_r d_{ij,r}) = 0 \tag{34}$$

Thus, we have a total of $(4M+4)$ moment conditions for $(2M+4)$ parameters, and the model is identified (over-identified). The MSM estimator minimizes the (simulated) generalized method of moments objective function, and we bootstrap to compute standard errors due to the small number of networks (Horowitz, 2001). Note that the above moment conditions are given in terms of pairs and/or individuals. With R independent networks, we specify moment conditions by aggregating the restrictions in (31) over pairs and the restrictions in (34) over individuals. Appendix C provides further details on how we construct the network-level moment conditions and on the MSM estimation procedure.

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).[13]

The AddHealth 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.

[13]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.

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).[14] 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 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. We symmetrize the social interactions by choosing the largest number between n_{ij} and n_{ji}. 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. The maximum geographical distance between two students, which is calculated for each network separately, is about 173 kilometers. The average distance is about 8.8 kilometers, while the median distance is about 5.3 kilometers.

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 about 2,236 students.[15] In addition, we focus on network sizes between 4 and 70 members.[16] We do this for two reasons. Firstly, the upper and lower tails of the distribution of networks by network size are commonly trimmed since the strength

[14]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.

[15]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 strenght of interactions is missing.

[16]In our data, the maximum network size is 68, the average one contains roughly 8 students, with a notable dispersion around this mean value (standard deviation equal to 10 pupils).

of peer effects may be too different in too small or too large networks (see Calvó-Armengol et al., 2009). Secondly, and most importantly, the computational burden of our estimation procedure requests to keep sample sizes relatively small. 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 4-70 agents. Table 1 shows that differences in means between these samples are almost never statistically significant. Our final sample consists of about 900 individuals distributed over roughly 100 networks. Among the adolescents selected in our sample of students, 58% are female and 20% are blacks. Slightly more than 70% live in a household with two married parents. The average parental education is high school graduate. The performance at school, as measured by the grade point average or GPA, exhibits a mean of 2.98, meaning slightly less than a grade of "B". The average family income is 44,562 in 1994 dollars, although 11% of parents chose not to report such information.

[*Insert Table 1 here*]

5.2 Empirical results

Table 2 displays the estimation results. In Columns $(1)-(2)$, we report the estimations when we use equation (22), i.e. we use the set of social distances where the dyadic relationship is *directional*, i.e. $(x_i - x_j)$. In Columns $(3)-(4)$, we use the set of social distances where the dyadic relationship is *undirectional*, i.e. $|x_i - x_j|$. Furthermore, in Columns (1) and (3), we use two basic sets of demographic variables, while Columns (2) and (4) extend the number of variables to include all other socio-demographic characteristics possibly related to the intensity of social interactions and social capital, such as household size.

Table 2 shows that the estimates are remarkably similar across Columns $(1)-(2)$ and across Columns $(3)-(4)$. Moreover, the estimates for the main structural parameters, or (n_0, α, c) do not differ substantially across all specifications. Therefore, our explanation of results will be mostly based on Columns (2) and (4) in which we use all available explanatory

variables.

[*Insert Table 2 here*]

Remember that the variable "Female" in Table 2 is a dummy variable that takes 1 if the respondent is a female (see Table 1). Let us therefore interpret Column (2) in Table 2 using equation (22) where $\beta_1 = -0.00033$ and $\beta_{M+1} = 0.0048$. This means that a pair of females is associated with 0.0096 more social interactions than a pair of males.[17] Furthermore, if a male and a female student meet ($x_i = 0$ and $x_j = 1$), the male student would have 0.00513 social interactions while the female student would have 0.00477 social interactions.[18] Given that the unit of social interactions is a weekly frequency, these estimates indicate that the difference is much less than one interaction per year, and hence economically insignificant. Table 3 summarizes our discussion by giving all the possible values of the effects of gender on social interactions using equation (22).

Table 3: Coefficients on $(x_i - x_j)$ and $(x_i + x_j)$

j (partner) / i (self)	Female	Male
Female	$2\beta_{M+1}$	$-\beta_1 + \beta_{M+1}$
Male	$\beta_1 + \beta_{M+1}$	0

The coefficients β_1 and β_{M+1} can be found in equation (22).

In particular, consider a student pair (i, j). A characteristic (e.g. female) of student j is associated with more social interactions if each item of the first column is larger than that of the second column in Table 3. Hence, the individual j is socially prefered as social interaction partner if $\beta_{M+1} > -\beta_1$. The opposite characteristic is preferred under the opposite condition. Applying this interpretation to the significant coefficient β_{M+1} in Column (2) of Table 2, we can see that students are preferred as social interaction partners if they are female, black, older students (higher grade), are physically more developed and are more religious. The same applies for students with better parental education, lower family income and less than two parents. Reciprocally, the characteristic of student i is associated with more interactions if each element of the first row is larger than that of the second row: that is, if $\beta_{M+1} > \beta_1$. Applying this criterion to the significant coefficient β_{M+1} in Column (2) of Table 2, all

[17] Indeed, $x_i = x_j = 1 \Rightarrow \beta_1(x_i - x_j) + \beta_{M+1}(x_i + x_j) = 0 + 2 \times 0.0048 = 0.0096$, and $x_i = x_j = 0 \Rightarrow \beta_1(x_i - x_j) + \beta_{M+1}(x_i + x_j) = 0$.

[18] Indeed, $\beta_1(x_i - x_j) + \beta_{M+1}(x_i + x_j) = 0.00033 + 0.0048 = 0.00513$ and $\beta_1(x_j - x_i) + \beta_{M+1}(x_j + x_i) = -0.00033 + 0.0048 = 0.00447$.

above reported characteristics apply except parental education and family composition. In these cases, a student with lower parental education and two parents tends to have more social interaction. Finally, a characteristic *unambiguously* yields more social interaction if the elements of its column and row are larger (that is, $2\beta_{M+1} > \max\{-\beta_1 + \beta_{M+1}, \beta_1 + \beta_{M+1}, 0\}$). Under this condition, we can see that students are unambiguously preferred as social interaction partners if they are female, black, good student (high grade), physically more developed and religious.

When we consider the symmetric social distances, i.e. Column (4), students' preferences exhibit homophily in all of their own characteristics if the coefficient β_{M+1} is negative and significantly different from zero. This occurs for female, black, grade, GPA, physical development, and religion practice. The estimates are all negative and significant, and their magnitudes are in general larger than those from the directed social distances in Column (2), which also supports homophily behaviors. When it comes to family background, students have homophily in family size and having two parents, but they do not have homophily in parental education and family income. The degree of homophily is the largest in student grade.

Next, the estimated coefficients on the $(x_i + x_j)$ variables exhibit mixed signs. The results in Columns (2) and (4) are mostly similar to each other but different for variables such as female, religion practice, and family income refused. In both specifications, the intensity of social interactions is increasing if a pair has a student with the following socio-demographic characteristics: black or other non-white race, higher grade, lower GPA, more physically developed, more parental education, less family income, two parents, and a smaller family size. These results are all statistically significant. It is particularly intuitive that older students have more social interactions because those who are 16 years old or older can drive to a friend's house. One possibly counter-intuitive result is that of family income. One would think students with more family income are likely to have more social interactions. However, it is possible that students from high-income families may be involved in more expensive types of activities such as playing sports (e.g. horseback riding, gymnastics, etc.), which is not distinguishable in our data. Alternatively, students from high-income families may be more likely to have best friends at other schools.

Turning our attention to the structural parameters, the baseline level of social interactions n_0 is roughly $2.2 - 3.4$, which is close to the optimal level of social interactions as we will see

in Section 6. The estimated cost of transportation is $0.00011 - 0.00026$ across specifications and statistically significant.

Combined with average pairwise distance (8.78 kilometers), the average estimated transportation cost is $0.001 - 0.0023$. Although the magnitude of all the estimates are small, the magnitudes of those estimates should be interpreted relative to each other. Given that the estimated cost of transportation per kilometer is only $0.00011 - 0.00026$, the magnitudes of homophily parameters ranging between 0.00022 and 0.0233 are not small.

6 Predictions and policy analysis

6.1 Dispersion

We now test Proposition 3 without using the structural estimates of our model. This proposition shows that a mean preserving increase in the spread of a symmetric distribution λ decreases social capital for all agents. To test this proposition, we can proceed as follows. For each of our 104 networks, we know the coordinates $(x_{i,r}, y_{i,r})$ of each individual i and the network baricenter $(\overline{x}_r, \overline{y}_r) = (\frac{1}{N_r}\sum_i x_{i,r}, \frac{1}{N_r}\sum_i y_{i,r})$. We then compute two types of geographical distance between network members: first, the average distance between individuals and their network baricenter, $\overline{d}_r = \frac{1}{N_r}\sum_i d_{i,r}$, and second, the average distance between pairs of members, $\tilde{d}_r = \frac{2}{N_r(N_r-1)}\sum_i \sum_{j\neq i} d_{ij,r}$. Furthermore, we calculate, for each network r, the average interactions $\overline{n}_r^*$ and the average social capital $\overline{s}_r^*$. Then, Proposition 3 predicts that there is *negative* relationship between $\overline{n}_r^*$ and $\overline{d}_r$ and between $\overline{s}_r^*$ and $\overline{d}_r$. We have the same prediction with $\tilde{d}_r$.

To verify these theoretical predictions, without claiming causality, we regress the average interactions $\overline{n}_r^*$ and the average social capital $\overline{s}_r^*$ on the average distance $\overline{d}_r$ as follows:

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

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

where N_r is the size (in terms of population) of network r, z_r are network measures and x_r are network-level socio-economic control variables (such as average family income in network r, etc.). Table 4 displays the results. It confirms that there is a *negative* and *significant* rela-

tionship between $\overline{n}_r^*$ and $\overline{d}_r$, and between $\overline{s}_r^*$ and $\overline{d}_r$. In terms of magnitude, a one kilometer increase in the geographical dispersion of individuals is associated with an approximately 0.06–0.07 decrease (5–6% decrease relative to the mean) in the average social interactions and a 0.002 decrease (0.2% decrease) in the average social capital. Such decreases are XXX What are these results relative to the mean (in percent terms?) The results are robust to the choice of the dispersion measures $\overline{d}_r$ and $\tilde{d}_r$. Note that we use the average, instead of the variance (or standard deviation) of pairwise distances because the latter captures the dispersion of distances rather than the dispersion of students' geographic locations. These results thus empirically confirm that distance is associated with lower levels of social interactions and social capital (see also Büchel and von Ehrlich (2017) for a similar result).

[*Insert Table 4 here*]

From Table 4, we also see that there is a non-monotonic relationship between the average interactions in a given network $\overline{n}_r^*$ and the network size N_r.[19] Using Column (4), we have:

$$\frac{\partial \overline{n}_r^*}{\partial N_r} = \gamma_1 + 2\gamma_2 N_r = 0.143 - 2(0.0016)N_r = 0 \tag{35}$$

Solving this equation leads to: $N_r = \frac{0.143}{2(0.0016)} \approx 44$. This means that the average social interactions increases with the network size until it reaches (approximately) 44 students and then decreases. $N_r = 44$ is thus the size of the network that *maximizes* average social interactions in our data. The result implies that students do not increase social interactions if they are connected with more than a certain number of friends.

6.2 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 perform simulations. We know from the theoretical model (Section 3) that, if the planner optimally chooses $n(x, y)$ and $s(y)$, we obtain:

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

[19] We do not comment on the relationship between the average social capital $\overline{s}_r^*$ and N_r since it is not significant.

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

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

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.

Given linear travel costs and the uniform distribution of individuals in a linear city, we have

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

$$2b\chi(y) = \int_{-b}^{b} \left\{ \left(n_0 + \theta(x,y)\right) n^o(x,y) - \frac{1}{2}\left[n^o(x,y)\right]^2 + \alpha\chi(x) n^o\left(x,y\right) \right\} \mathrm{d}x$$

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

Let us now determine the equations we want to calibrate. We have:

$$n^o_{ij,r} = n_0 - \frac{c d_{ij,r}}{s^o_{j,r}} + \alpha\chi_{i,r}\, s^o_{j,r} + \theta_{ij,r}, \tag{36}$$

$$2b_r\chi_{j,r} = \sum_{i=1}^{Nr} \left\{ \left(n_0 + \theta_{ij,r}\right) n^o_{ij,r} - \frac{1}{2}\left(n^o_{ij,r}\right)^2 + \alpha\chi_{i,r}\, n^o_{ij,r} \right\}, \tag{37}$$

$$s^o_{j,r} = 1 + \frac{\alpha}{2b_r} \sum_{k=1}^{N} n^o_{jk,r} s^o_{k,r} \tag{38}$$

Here is how we proceed. From the previous estimations of the equilibrium model, we have the estimated values of n_0, α, c and $\theta_{ij,r}$. From the data, we know b_r and $d_{ij,r}$. By plugging these values into (36), (37) and (38), 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 $\upsilon_{i,r}$ 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 $2N_r + L_r$ equations since there are L_r equations for (36), N_r equations for (37) and N_r equations for (38).

When we have calculated all the $n^o_{ij,r}$, $s^o_{j,r}$ and $\chi_{i,r}$, we can then compare the observed values of $n^*_{ij,r}$ in equilibrium and at the social (first best) optimum $n^o_{ij,r}$. We can also compare the predicted value of $s^*_{j,r}$, which is calculated by using equation (28) with our parameter

estimates and at the first best $s^o_{j,r}$ (see equation (38)). According to Proposition 5, we should find that students socially interact too little compared to the social optimal, i.e. $n^o_{ij,r} > n^*_{ij,r}$, $\forall i, j$, and $s^o_{j,r} > s^*_{j,r}$, $\forall i_{xr}$.

We numerically solve the optimal level of social interactions and social capital with the MSM parameter estimates in Column (6) of Table 2. Table 5 displays the results. On average, each pair interacts 1.2 fewer times than is socially optimal. The difference between the socially optimal and the observed levels of social interactions varies from -0.81 to 2.74 across networks. Although there are a few networks where the observed level is larger than the optimal level, most networks' interactions fall short of the optimum. Students also have less social capital than optimal (by 0.236, or approximately 25%, on average). Note that our measure of social capital is based on data augmentation using the calibrating equation (20). We compare this augmented social capital with the optimal one computed from equations (36)–(38).

[*Insert Table 5 here*]

Network size and social interactions Furthermore, we would like to find which variables are closely associated with the discrepancy between the optimal level and the observed level. To see this, 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:

$$\overline{n}^o_r - \overline{n}^*_r = \gamma_0 + \gamma_1 N_r + \gamma_2 \left(N_r\right)^2 + \gamma_3 \overline{d}_r + \gamma_z z_r + \gamma_x x_r + \epsilon_r, \tag{39}$$

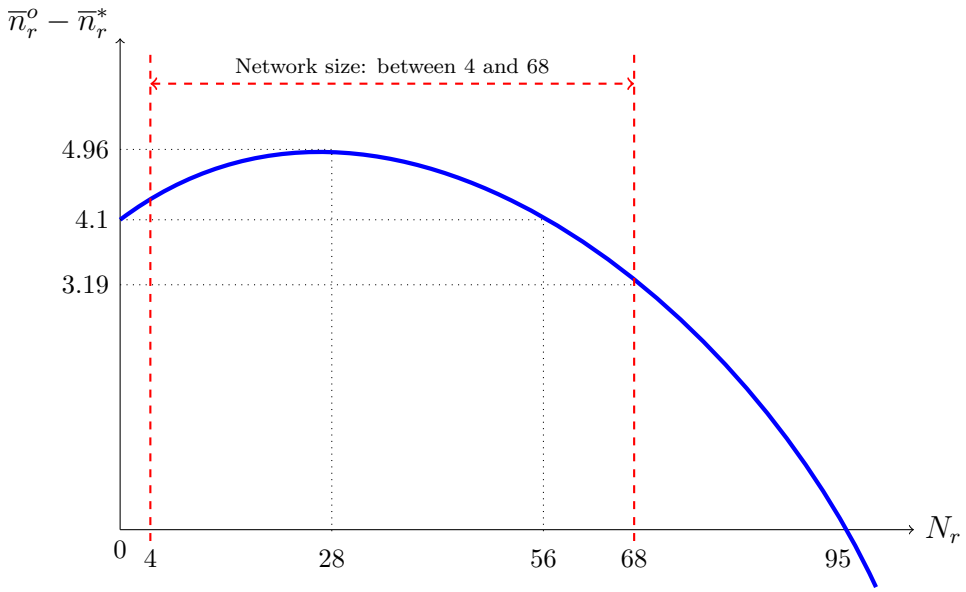
$$\overline{s}^o_r - \overline{s}^*_r = \delta_0 + \delta_1 N_r + \delta_2 \left(N_r\right)^2 + \delta_3 \overline{d}_r + \delta_z z_r + \delta_x x_r + \zeta_r. \tag{40}$$

Tables 6–7 show the results. Consider first *social interactions* (Table 6) 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:

$$\frac{\partial(\overline{n}^o_r - \overline{n}^*_r)}{\partial N_r} = \gamma_0 + 2\gamma_2 N_r = 0.0614 - 2(0.0011) N_r = 0 \tag{41}$$

Solving this equation leads to: $N_r = \frac{0.0614}{2(0.0011)} \approx 28$. This means that the difference between the optimal and the observed level of social interactions is increasing until the network

Figure 1: Difference between optimal and observed social interactions by network size



size reaches (approximately) 28 students and then decreases. As a result, there is a non-monotonic relationship between $\overline{n}_r^o - \overline{n}_r^*$ and N_r where an increase in the network size increases $\overline{n}_r^o - \overline{n}_r^*$ up to $N_r = 28$ and, above this size, an increase in the network size decreases $\overline{n}_r^o - \overline{n}_r^*$. Thus, $N_r = 28$ is the size of the network that *maximizes* these inefficiencies. Given that the average size of the networks is $N_r^{av} = 8$, in terms of magnitude, an increase by one person in the network from $N_r^{av} = 8$, raises these inefficiencies by $0.0614 - 2(0.0011)N_r^{av} = 0.0438$. Figure 1 illustrates how the difference changes along with the network size N_r.

Let us now determine the network size that *minimizes* these inefficiencies. Since there is a hump-shaped relationship between $\overline{n}_r^o - \overline{n}_r^*$ and N_r (Figure 1), it has to be either at N_r^{min} or at N_r^{max}, the minimum and maximum network size. In our data (Section 5.1), we have: $N_r^{min} = 4$ and $N_r^{max} = 68$. We test this equation, and Figure 1 shows that $N_r^{max} = 68$ *minimizes* the inefficiencies in terms of social interactions.

Furthermore, among the network measures, in Table 6, we see that average degree, the average eigenvector centrality, and the network diameter are negatively correlated with $\overline{n}_r^o - \overline{n}_r^*$. This means, for example, that the more "spread" is the network (in terms of diameter), the lower are the inefficiencies in social interactions. Moreover, the average characteristics of the students are also associated with the optimal-observed difference in social interactions. Networks that consist of students with higher grades, lower GPA, smaller family size, and lower family income are more likely to have higher inefficiencies in terms of social interactions.

In particular, it is likely that networks of students from a low-income family have higher inefficiencies due to financial pressure.

Let us now turn to the inefficiencies in terms of social capital (Table 7). We have seen, previously, in Table 5, that these inefficiencies were very small. As a result, quite naturally, in Table 7, we see that the effect of the different variables on $(\overline{s}_r^o - \overline{s}_r^*)$ are nearly never large and significant. We only find a weak negative relationship between the average geographic distance and $\overline{s}_r^o - \overline{s}_r^*$.

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.

[*Insert Tables 6 and 7 here*]

Network size and average welfare Another interesting exercise, for which do not have theory, is to determine the optimal network (i.e. the one that maximizes total welfare).[20] 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 104 networks. Remember that the welfare per network is given by:

$$W_r^* = \frac{1}{4b_r^2} \sum_{i=1}^{N_r} \sum_{j=1}^{N_r} \left[\left((n_0 + \theta_{ij,r})\, n_{ij,r}^* - \frac{1}{2} \left(n_{ij,r}^* \right)^2 \right) s_{j,r}^* - n_{ij,r}^* c d_{ij,r} \right] \tag{42}$$

As a result, the average welfare per network is:

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

We would like know which network r maximizes AW_r^*, i.e. $\max_r AW_r^*$.

[20] Determining the optimal network is a very difficult exercise; see König et al. (2014) and Belhaj et al. (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.

For that, we run the following regression:

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

to see 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.[21] We include the network measures to see how the shape of a network is associated with the welfare.

Table 8 reports the results. We can first calculate the network size that maximizes the average welfare per network AW_r^*. Using Column (5), we have:

$$\frac{\partial AW_r^*}{\partial N_r} = \delta_1 + 2\delta_2 N_r = 2.071 - 2(0.0187)N_r = 0 \tag{43}$$

Solving this equation leads to: $N_r = \frac{2.071}{2(0.0187)} \approx 55$. This means the network that comprises (approximately) 55 students is the one that maximizes the average welfare per network. This is a network of a big size given that the largest network has 68 students. We have, however, to be careful with the interpretation of this result since the effect of N_r on AW_r^* is non-significant (Table 8).

In Table 8, we also find that the average pairwise geographic distance is an important factor for designing an optimal network. The higher the distance between two individuals in a network, the lower the average welfare. This result is closely related to our previous finding of the negative relationship between social interactions and dispersion in Section 6.1 (Table 4). What is interesting is that the results of Tables 4 and 8 are similar, even though the latter are obtained using the structural estimates of our model while the former are directly derived from the data without using any structural estimation. This gives us confidence that our model fits the data well.

[*Insert Table 8 here*]

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

6.3 Policies

Once we have estimated this model, we can implement the two policies suggested in Proposition 6. The first policy consists in *subsidizing social interactions*. The second policy consists in *subsidizing the transport cost* per individual c. We then evaluate their impact on $n(x,y)$, the frequency of interactions. Which policy is more effective at moving the observed interactions/social capital closer to the optimal levels? We have seen in Proposition 6 that the first best solution can be restored if social interactions are not subsidized while commuting trips are subsidized as a function of the locations of the des- tination and origin partners. The latter is unlikely to be implemented and this is why we now consider *social-interaction subsidies and travel-cost subsidies that only target each individual but not a pair of individuals*.

6.3.1 Subsidizing social interactions

We assume that the planner subsidizes the intensity of social interactions $n(x,y)$ in the following way:

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

where σ is the value of the social-interaction subsidy. In this formulation, each individual x receives a fixed amount of money $\sigma \int_{-b}^{b} n(x,y)\lambda(y)\mathrm{d}y$ proportional to the individual x's social interaction effort with all her friends, i.e. $\int_{-b}^{b} n(x,y)\lambda(y)\mathrm{d}y$. The government (or the planner) is here introduced as an agent that can set subsidy rates on social interaction effort before the individuals decide upon their social interaction 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.[22] One example of such a subsidy is to support social mixing by providing a community center for students' activities.

Given linear travel costs and the uniform distribution of individuals in a linear city, we have

$$U(x) = \frac{1}{2b} \int_{-b}^{b} \left[\left(\left(n_0 + \theta(x,y)\right) n\left(x,y\right) - \frac{1}{2}\left[n\left(x,y\right)\right]^2 \right) s(y) - \left(c\left|x-y\right| - \sigma\right) n(x,y) \right] \mathrm{d}y$$

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

In that case, it is easily verified that the equilibrium is now given by:

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

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

where the superscript σ on variables is used to denote the subsidy policy.

From the data, we would like to consider these two equations written as follows:

$$n^{\sigma}_{ij,r} = n_0 + \frac{\sigma_r - cd_{ij,r}}{s^{\sigma}_{j,r}} + \theta_{ij,r} \tag{44}$$

and

$$s^{\sigma}_{j,r} = 1 + \frac{\alpha}{2b_r}\sum_{k=1}^{N_r} n^{\sigma}_{jk,r}\, s^{\sigma}_{k,r} \tag{45}$$

The total welfare per network is *now* defined as

$$\begin{aligned}
W &= \int_{-b}^{b} U(x)\lambda(x)\mathrm{d}x \\
&= \int_{-b}^{b}\int_{-b}^{b} \left\{v\left(n(x,y)\right)s(y) - n(x,y)c(x-y)\right\}\lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y + \sigma\int_{-b}^{b}\int_{-b}^{b} n(x,y)\lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y \\
&= \frac{1}{4b^2}\int_{-b}^{b}\int_{-b}^{b}\left[\left(\left(n_0+\theta(x,y)\right)n\left(x,y\right) - \frac{1}{2}\left[n\left(x,y\right)\right]^2\right)s(y) - n(x,y)c\,|x-y|\right]\mathrm{d}x\mathrm{d}y \\
&\quad + \frac{\sigma}{4b^2}\int_{-b}^{b}\int_{-b}^{b} n(x,y)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

For the estimation, the total welfare per network is equal to

$$W^{\sigma}_{r} = \frac{1}{4b_r^2}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\left[\left(\left(n_0+\theta_{ij,r}\right)n^{\sigma}_{ij,r} - \frac{1}{2}\left(n^{\sigma}_{ij,r}\right)^2\right)s^{\sigma}_{j,r} - \left(cd_{ij,r} - \sigma_r\right)n^{\sigma}_{ij,r}\right] \tag{46}$$

We find the subsidy σ_r^* that gives network r the same aggregate utility W_r^{σ} as the first best W_r^o. From the estimated value of the equilibrium model, we have α, c and n_0; from the data we have $d_{ij,r}$ and b_r. We can then numerically solve equations (44) and (45) and the optimal subsidy that maximizes (46) to obtain σ_r^*, $n^{\sigma}_{ij,r}$ and $s^{\sigma}_{j,r}$. See Appendix D for technical details.

The first three columns in Table 9 display the result. On average, a subsidy level of 0.4133 for each social interaction is required for a network to achieve the first best aggregate

level of social interactions and social capital. Most networks are offered a positive subsidy, which reflects a lack of social interaction. Nevertheless, negative subsidies are given to three networks, which empirically have interaction levels above the optimum. We also compute a single subsidy σ^* for all networks, which allows individuals to achieve the first best as close as possible and we find $\sigma^* = 1.4534$.

[*Insert Table 9 here*]

6.3.2 Subsidizing transportation costs

In the theoretical model, each agent paid a marginal transport cost per distance equal to c. Now, it is given by $c - \tau$ so that τ is a subsidy on transportation costs financed by the government. In that case, the total social interaction cost of an agent located at x is now given by

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

so that it is less costly to commute for interacting with other agents. In this case, the equilibrium equations are defined as:

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

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

It is clear from the theory that an increase in τ increases the levels of both social interactions $n^{\tau}(x,y)$ and social capital $s^{\tau}(x)$ for each agent.

From the data, we would like to consider these two equations written as follows:

$$n^{\tau}_{ij,r} = n_0 - \frac{(c-\tau_r) d_{ij,r}}{s^{\tau}_{j,r}} + \theta_{ij,r} \tag{47}$$

$$s^{\tau}_{j,r} = 1 + \frac{\alpha}{2b_r} \sum_{k=1}^{N_r} n^{\tau}_{jk,r} \, s^{\tau}_{k,r} \tag{48}$$

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

$$W_r^\tau = \frac{1}{4b_r^2} \sum_{i=1}^{N_r} \sum_{j=1}^{N_r} \left[\left((n_0 + \theta_{ij,r})\, n_{ij,r}^\tau - \frac{1}{2} \left(n_{ij,r}^\tau \right)^2 \right) s_{j,r}^\tau - n_{ij,r}^\tau \,(c - \tau_r)\, d_{ij,r} \right] \tag{49}$$

As for the social interaction subsidy, we find the subsidy τ_r^* that gives the same aggregate utility W_r^τ in network r as the first best W_r^0. From the estimated value of the equilibrium model, we have α, c and $n_{0,r}$, and from the data $d_{ij,r}$ and b_r. We can then numerically solve equations (47) and (48) and the optimal subsidy that maximizes (49) to obtain τ_r^*, $n_{ij,r}^\tau$ and $s_{j,r}^\tau$.

The last three columns in Table 9 display the results. On average, a subsidy level of 0.9471 per kilometer is required for a network to achieve the first best aggregate level of social interactions and social capital. As above, most networks receive positive subsidies to entice more interactions. We also compute a single subsidy τ^* for all networks, which allows individuals to achieve the first best as close as possible, and we find $\tau^* = 5.8601$.

6.3.3 Comparing the two policies

Finally, it is interesting to compare these two policies at a given cost. The question is then as follows: Given that the planner has an amount T to spend, which policy should she choose? In order to distribute a total amount of subsidy T to each network, we consider three different schemes. First, we distribute the same amount $T_r = T/R$ for each network (uniform subsidy). The second scheme gives an amount proportional to network population N_r. Hence, $T_r = \frac{N_r}{\sum_{r'} N_{r'}} T$. The last subsidy scheme provides an amount proportional to the number of pairs $N_r(N_r - 1)$, i.e. $T_r = \frac{N_r(N_r-1)}{\sum_{r'} N_r'(N_{r'}-1)} T$.

Let us thus write the budget constraint for each policy. For the *social-interaction subsidy policy*, the planner's budget constraint for each network r can be written as:

$$\frac{\sigma_r}{4b_r^2} \sum_{i=1}^{N_r} \sum_{j=1}^{N_r} n_{ij,r} = T_r \tag{50}$$

where the left-hand side is the total cost of the policy for each network and T is the fixed amount that needs to be spent. For the *transportation subsidy policy*, the planner's budget

constraint for each network can be written as:

$$\frac{\tau_r}{4b_r^2}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r} n_{ij,r}d_{ij,r} = T_r \tag{51}$$

We shall proceed as follows. First, we consider the *social-interaction subsidy policy.* We observe $d_{ij,r}$ and b_r in the data and have estimated α, c and n_0. We fix T to a fixed value (say 66,000) and $R = 104$, and can then solve simultaneously equations (44), (45) and (50). We get the different endogenous variables, in particular, the different subsidies σ_r. For this each value of σ_r, we calculate the total welfare W_r^σ given by (46). Then, we can calculate TW^σ, the total welfare in the economy, i.e. $TW^\sigma = \sum_{r=1}^R W_r^\sigma$, or equivalently

$$\begin{aligned} TW^\sigma =& \sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\frac{1}{4b_r^2}\left(\left(n_0+\theta_{ij,r}\right)n_{ij,r}^\sigma - \frac{1}{2}\left(n_{ij,r}^\sigma\right)^2\right)s_{j,r}^\sigma \\ &- \sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\frac{1}{4b_r^2}n_{ij,r}^\sigma c\, d_{ij,r} + \sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\frac{\sigma_r}{4b_r^2}n_{ij,r}^\sigma \end{aligned} \tag{52}$$

Second, we consider the *transportation subsidy policy.* We observe $d_{ij,r}$ and b_r in the data and have estimated α, c and n_0. We fix T to a fixed value and $R = 104$, and can then solve simultaneously equations (47), (48) and (51). We get the different endogenous variables, in particular, the different subsidies τ_r. Then, for each value of τ_r, we can calculate the total welfare W_r^τ given by (49). Then, we can calculate TW^τ, the total welfare in the economy, i.e. $TW^\tau = \sum_{r=1}^R W_r^\tau$, or equivalently

$$TW^\tau = \frac{1}{4b^2}\sum_{r=1}^{R}\sum_{i=1}^{N_r}\sum_{j=1}^{N_r}\left[\left(\left(n_0+\theta_{ij,r}\right)n_{ij,r}^\tau - \frac{1}{2}\left(n_{ij,r}^\tau\right)^2\right)s_{j,r}^\tau - \frac{1}{4b^2}n_{ij,r}^\tau\left(c-\tau_r\right)d_{ij,r}\right] \tag{53}$$

The key question is then whether $TW^\tau \gtreqless TW^\sigma$.

Table 10 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. In particular, the social-interaction policy is more effective for most networks under the two first schemes. Under the subsidy scheme proportional to the number of pairs, the performance of

the transportation policy is relatively closer to the social-interaction policy but still lower.[23] As a result, if a planner has a given amount of money to spend, she should subsidize social interactions and not transportation costs because it yields greater improvements to total welfare.[24]

[*Insert Table 10 here*]

7 Concluding remarks and policy implications

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 is an interesting result, which seems to be confirmed by what we observe in world-world cities. Indeed, despite rapid innovation in communication technologies, we still observe an important growth in urbanization, which may highlight the importance of geographical proximity for social exchange (see, e.g. Goldenberg and Levy (2009)). 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

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

[24] Observe that this result is not in contradiction with Proposition 6, which shows that to restore the first-best solutions social interactions should not be subsidized while transportation costs should be. In Table 10, 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 9, we are calculating the subsidy level that restores the first best for each policy. We see that, in order to restore the first best, transportation costs should be much more subsidized than social interactions.

there is a non-monotonic relationship between the inefficiencies in terms of social interactions and the network size. In fact, these inefficiencies are the largest when the network is composed of 28 students and the smallest for 68 students. On the contrary, we find that the network that maximizes the average welfare in a network should have 55 students whereas the one that maximizes social interactions should be of size 44. There is therefore a discrepancy between maximizing average welfare, maximizing social interactions and minimizing the inefficiencies of social interactions. We then perform two different subsidy policies. Our results suggest that the individuals interact at optimal levels when 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.[25] 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 et al. (2001), Kling et al. (2007) and Chetty et al. (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. (2018) 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)[26] but also because of the direct effects social interactions

[25] 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.

[26] 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

have on innovation (Bailey et al. (2017)) 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.

References

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

Badev, A. (2017). Discrete games in endogenous networks: Theory and policy. Technical report, Mimeo., University of Pennsylvania.

Bailey, M., Cao, R. R., Kuchler, T., and Stroebel, J. (2018). The economic effects of social networks: Evidence from the housing market. *Journal of Political Economy*, forthcoming.

Bailey, M., Cao, R. R., Kuchler, T., Stroebel, J., and Wong, A. (2017). Measuring social connectedness. Technical report, National Bureau of Economic Research.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144):1236498.

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

Beaman, L. (2016). Social networks and the labor market. In Bramoulé, Y., Galeotti, A., and Rogers, B., editors, *Oxford Handbook on the Economics of Networks*. Oxford University Press, Oxford.

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

Behrens, K., Duranton, G., and Robert-Nicoud, F. (2014). Productive cities: Sorting, selection, and agglomeration. *Journal of Political Economy*, 122(3):507–553.

the importance of these intellectual spillovers (Combes and Gobillon (2015)).

Belhaj, M., Bervoets, S., and Deroïan, F. (2016). Efficient networks in games with local complementarities. *Theoretical Economics*, 11(1):357–380.

Berliant, M., Peng, S.-K., and Wang, P. (2002). Production externalities and urban configuration. *Journal of Economic Theory*, 104(2):275–303.

Berry, S. (1992). Estimation of a model of entry in the airline industry. *Econometrica*, 60(4):889–917.

Bifulco, R., Fletcher, J. M., Ross, S. L., et al. (2011). The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health. *American Economic Journal: Economic Policy*, 3(1):25–53.

Bisztray, M., Koren, M., and Szeidl, A. (2017). Learning to import from your peers. Technical report, Central European University,.

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

Büchel, K. and von Ehrlich, M. (2017). Cities and the structure of social interactions: Evidence from mobile phone data. Technical report, CESifo Working Paper Series No. 6568.

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

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

Card, D. and Giuliano, L. (2013). Peer effects and multiple equilibria in the risky behaviors of friends. *Review of Economics and Statistics*, 95(4):1130–1149.

Chetty, R., Hendren, N., and Katz, L. F. (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.

Ciliberto, F. and Tamer, E. (2009). Market structure and multiple equilibria in airline markets. *Econometrica*, 77(6):1791–1828.

Combes, P.-P. and Gobillon, L. (2015). The empirics of agglomeration economies. In Duranton, G., Henderson, V., and Strange, W., editors, *Handbook of Regional and Urban Economics, Volume 5A*, pages 247–348. Elsevier, Amsterdam.

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

Currarini, S., Jackson, M. O., and Pin, P. (2010). Identifying the roles of race-based choice and chance in high school friendship network formation,. *Proceedings of the National Academy of Science of the USA*, 107(11):4857–4861.

Duranton, G. and Puga, D. (2015). Urban land use. *Handbook of Regional and Urban Economics*, 5:467–560.

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

Fujita, M. (1989). *Urban economic theory: land use and city size*. Cambridge university press.

Fujita, M., Krugman, P. R., Venables, A. J., and Fujita, M. (1999). *The spatial economy: cities, regions and international trade*, volume 213. Wiley Online Library.

Fujita, M. and Thisse, J.-F. (2013). *Economics of agglomeration: cities, industrial location, and globalization*. Cambridge university press.

Geyer, J. (2017). Housing demand and neighborhood choice with housing vouchers. *Journal of Urban Economics*, 99:48–61.

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

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

Glaeser, E. L. and Sacerdote, B. (2000). The social consequences of housing. *Journal of Housing Economics*, 9:1–23.

Goldenberg, J. and Levy, M. (2009). Distance is not dead: Social interaction and geographical distance in the internet era. *Computers and Society*, 2:1–22.

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

Hellerstein, J. K., Kutzbach, M. J., and Neumark, D. (2014). Do labor market networks have an important spatial dimension? *Journal of Urban Economics*, 79:39–58.

Hellerstein, J. K., McInerney, M., and Neumark, D. (2011). Neighbors and coworkers: The importance of residential labor market networks. *Journal of Labor Economics*, 29(4):659–695.

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

Helsley, R. W. and Strange, W. C. (2014). Coagglomeration, clusters, and the scale and composition of cities. *Journal of Political Economy*, 122(5):1064–1093.

Helsley, R. W. and Zenou, Y. (2014). Social networks and interactions in cities. *Journal of Economic Theory*, 150:426–466.

Horowitz, J. L. (2001). The bootstrap. *Handbook of econometrics*, 5:3159–3228.

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

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

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

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

Jackson, M. O., Rogers, B., and Zenou, Y. (2017). The economic consequences of social network structure. *Journal of Economic Literature*, 55(1):1–47.

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

Jackson, M. O. and Zenou, Y. (2015). Games on networks. *Handbook of Game Theory*, 4:91–157.

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

Katz, L., Kling, J., and Liebman, J. (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *Quarterly Journal of Economics*, 116:607—654.

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

König, M., Liu, X., and Zenou, Y. (2014). R&D networks: Theory, empirics and policy implications.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics*, 28(4):825–860.

List, J., Momeni, F., and Zenou, Y. (2017). Spillover effects in education. Technical report, University of Chicago.

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

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

Marmaros, D., Sacerdote, B., et al. (2006). How do friendships form? *The Quarterly Journal of Economics*, 121(1):79–119.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5):995–1026.

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

Mossay, P., Picard, P., et al. (2013). Spatial segregation and urban structure. Technical report, CORE Discussion Paper No. 2013/19.

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

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

Pakes, A. and Pollard, D. (1989). Simulation and the asymptotics of optimization estimators. *Econometrica*, 57(5):1027–1057.

Pischke, J.-S. (1995). Measurement error and earnings dynamics: Some estimates from the psid validation study. *Journal of Business & Economic Statistics*, 13(3):305–314.

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.

Romer, P. M. (1986). Increasing returns and long-run growth. *The Journal of Political Economy*, 94(5):1002–1037.

Rosenthal, S. S. and Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 46(2):373–389.

Sato, Y. and Zenou, Y. (2015). How urbanization affect employment and social interactions. *European Economic Review*, 75:131–155.

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

Sheng, S. (2015). Identification and estimation of network formation games. *Unpublished Manuscript*.

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

Zenou, Y. (2009). *Urban Labor Economics*. Cambridge University Press.

Zenou, Y. (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:

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

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

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

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 a mean preserving spread of the distribution of $\lambda(x)$. ∎

Linear travel costs and uniform distribution of agents: Proofs of (13) and (14)

Let us first calculate $g(y)$, which is given by:

$$g(y) = c_1 \int_{-b}^{y} (y - z) \lambda(z) \mathrm{d}z + c_1 \int_{y}^{b} (z - y) \lambda(z) \mathrm{d}z$$

We have:

$$\begin{aligned} g(y) &= c_1 \int_{-b}^{y} (y - z) \lambda(z) \mathrm{d}z + c_1 \int_{y}^{b} (z - y) \lambda(z) \mathrm{d}z \\ &= \frac{c_1}{2b} \left[\int_{-b}^{y} y \mathrm{d}z - \int_{-b}^{y} z \mathrm{d}z + \int_{y}^{b} z \mathrm{d}z - \int_{y}^{b} y \mathrm{d}z \right] \\ &= \frac{c_1 \left(y^2 + b^2 \right)}{2b} \end{aligned}$$

Let us now calculate $n^*(x, y)$ and $s^*(y)$. From Proposition 1, we obtain (13), that is:

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

and

$$s^*(y) = \frac{1-\alpha^2\int_{-b}^{b} g(z)\lambda(z)\mathrm{d}z}{1-\alpha} - \alpha g(z)$$
$$= \frac{1-\frac{\alpha^2}{2b}\int_{-b}^{b}\frac{c_1\left(z^2+b^2\right)}{2b}\mathrm{d}z}{1-\alpha} - \frac{\alpha c_1\left(z^2+b^2\right)}{2b}$$

It is easily verified that

$$\int_{-b}^{b}\frac{c_1\left(z^2+b^2\right)}{2b}\mathrm{d}z = \frac{c_1}{2b}\left(\frac{2}{3}b^3+2b^3\right) = \frac{4c_1b^2}{3}$$

Therefore,

$$s^*(y) = \frac{3-2\alpha^2c_1b}{3\left(1-\alpha\right)} - \frac{\alpha c_1\left(z^2+b^2\right)}{2b}$$

Alternatively, from (2), we obtain (14), that is:

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

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

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

Finally, we note that $\int_{-b}^{b}\chi(x)\left[s(x)-1\right]\lambda(x)\mathrm{d}x = \int_{-b}^{b}\chi(y)\left[s(y)-1\right]\lambda(y)\mathrm{d}y$. 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:

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

We now use variation calculus on the Lagragian 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 $\int_{-b}^{b} \{F'_s[n^o(x, y), s^o(y), x, y]\} \lambda(x)\mathrm{d}x = 0$. This gives (16) and (17). ∎

Proof of Proposition 5: Condition (16) yields

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

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 (9) 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 (7), 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 Correlation of unobserved variables

In this appendix, we allow for correlation of unobservables across pairs by incorporating individual-specific unobserved variables. Ignoring such clustering may lead to smaller standard errors and overstating significance of the results. Recall equation (21).

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

We let $\varepsilon_{ij,r} = \eta_{i,r} + \eta_{j,r} + \upsilon_{ij,r}$, where $\eta_{i,r}$ is individual-specific unobserved variable of i at network r, and $\upsilon_{ij,r}$ is a pair-specific unobserved variable. The individual specific unobserved variable $\eta_{i,r}$ is *i.i.d.* across i and r with mean zero and variance σ_η^2, and the pair-specific unobserved variable is *i.i.d.* across individuals, pairs, and networks with mean zero and variance σ_υ^2.

Under the above specification, each component in the variance-covariance matrix $\Omega_{\varepsilon,r}$ of the $N_r^2 \times 1$ vector of unobservables ε_r has the following form:

$$Cov(\varepsilon_{ij,r}, \varepsilon_{kh,r}) = \begin{cases} 2\sigma_\eta^2 + \sigma_\upsilon^2, & \text{if } i = k \text{ and } j = h. \\ 2\sigma_\eta^2, & \text{if } i = h \text{ and } j = k. \\ \sigma_\eta^2, & \text{if } ij \text{ and } kh \text{ share only one index.} \\ 0, & \text{otherwise.} \end{cases}$$

Table B1 shows the estimation results with correlated errors.

Our parameter estimates are close to the result in Table 2, which are obtained under the assumption of independently distributed pair-level unobserved variables. By incorporating clusters, standard errors becomes larger, but the increments are small in magnitudes. Hence, the significance levels of the estimation results are almost identical to those in Table 2. The estimated standard deviation σ_η of individual-specific unobserved variable is small as well.

Table B1: Structural estimation results with individual-specific unobserved variables

	(1)	(2)	(3)	(4)
**** Social interaction equation ****				
β: Social distances	$(x_i - x_j)$	$(x_i - x_j)$	$\lvert x_i - x_j \rvert$	$\lvert x_i - x_j \rvert$
Female	-0.00061**	-0.00033**	-0.00029**	-0.00022**
	(0.00031)	(0.00014)	(0.00012)	(0.00009)
Black	-0.0025**	-0.0019**	-0.0191**	-0.0127**
	(0.0011)	(0.0008)	(0.0090)	(0.0053)
Grade	-0.0046**	-0.0035**	-0.0321**	-0.0235**
	(0.0021)	(0.0015)	(0.0142)	(0.0093)
GPA	-0.0033**	-0.0037**	-0.0017**	-0.0017**
	(0.0014)	(0.0016)	(0.0007)	(0.0007)
Parental education	0.0118**	0.0078**	-0.0010**	0.0035**
	(0.0051)	(0.0032)	(0.0004)	(0.0015)
Family income	-0.0026**	-0.0021**	0.0011**	0.0012**
	(0.0012)	(0.0009)	(0.0005)	(0.0005)
Two parents	0.00059**	0.00035**	-0.0048**	-0.0033**
	(0.00025)	(0.00015)	(0.0022)	(0.0014)
Other race		-0.0145**		-0.0021***
		(0.0062)		(0.0008)
Physical development		0.0041**		-0.0031**
		(0.0017)		(0.0013)
Religion practice		-0.0039**		-0.0191**
		(0.0017)		(0.0079)
Family size		0.0041**		-0.0011**
		(0.0016)		(0.0005)
Family income refused		0.0031**		-0.0058**
		(0.0016)		(0.0023)
β: combined levels	$(x_i + x_j)$	$(x_i + x_j)$	$(x_i + x_j)$	$(x_i + x_j)$
Female	0.0040**	0.0048**	-0.0027**	-0.0030**
	(0.0017)	(0.0019)	(0.0011)	(0.0013)
Black	0.0212**	0.0207**	0.0086**	0.0062**
	(0.0094)	(0.0086)	(0.0042)	(0.0026)
Grade	-0.0006***	0.0057***	-0.0037***	0.0378***
	(0.00018)	(0.0021)	(0.0011)	(0.0132)
GPA	-0.00068***	-0.0117***	0.0008***	-0.0294**
	(0.00024)	(0.0044)	(0.00027)	(0.0131)
Parental education	0.00029***	0.00316***	0.00009***	0.00092**
	(0.0001)	(0.00119)	(0.00003)	(0.0004)
Family income	-0.0047**	-0.0038**	-0.0059**	-0.0049**
	(0.0020)	(0.0016)	(0.0026)	(0.0021)
Two parents	0.0050**	0.0053**	0.00009**	0.00008**
	(0.0021)	(0.0022)	(0.00004)	(0.00004)
Other race		0.0062**		0.0098**
		(0.0025)		(0.0038)
Physical development		0.0043**		0.0017**
		(0.0017)		(0.0007)
Religion practice		0.0054**		-0.0051**
		(0.0022)		(0.0022)
Family size		-0.0048**		-0.0098**
		(0.0020)		(0.0043)
Family income refused		0.0066**		-0.0014**
		(0.0027)		(0.0005)
n_0	2.9846***	2.4530**	2.9555***	2.2793***
	(1.0981)	(0.9987)	(1.1247)	(0.8839)
c (transportation cost)	0.00015**	0.00011**	0.00043*	0.00023**
	(0.00007)	(0.00005)	(0.00022)	(0.00010)
σ_η	0.0565**	0.0507**	0.0905**	0.0909**
	(0.0252)	(0.0225)	(0.0389)	(0.0366)
σ_υ	3.7179**	3.7147**	4.5468**	2.4079***
	(1.8271)	(1.5647)	(1.8191)	(0.9167)
**** Social capital equation ****				
α	0.0299**	0.0241**	0.0238**	0.0178**
	(0.0124)	(0.0099)	(0.0109)	(0.0078)
Number of networks	104	104	104	104
Number of pupils	890	890	890	890
Number of directed pairs	18,482	18,482	18,482	18,482

We estimate parameters $(n_0, c, \beta^{\mathrm{T}}, \sigma_\varepsilon)^{\mathrm{T}}$ in the social interaction equation (19) and subsequent specifications (22)–(21), and parameter α in the social capital equation (20).
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

C Estimation: technical details

C.1 Moment construction

We estimate the parameter vector $\theta = (n_0, \alpha, c, \beta^{\mathrm{T}}, \sigma_\varepsilon^2)^{\mathrm{T}}$ using MSM. We assume that the true parameter vector θ_0 lies in the interior of the compact parameter space $\Theta \subset \mathbb{R}^{2M+2}$, where M is the number of individual-level covariates. In our main specification, we use $M = 12$ variables.

Recall the prediction errors $\hat{\nu}_{ij,r}$ and $\hat{\xi}_{j,r}$ in equations (30) and (33). These prediction errors are mean independent of exogenous variables $x_{ij,r}$ and $d_{ij,r}$. Note that the first prediction error $\hat{\nu}_{ij,r}$ is pairwise, while $\hat{\xi}_{j,r}$ is individual. However, this discrepancy does not raise an issue, because the asymptotic argument in our MSM estimation is R, i.e. the number of networks. For the sake of exposition, let $x_{ij,r}$ subsume the constant, and hence, it is a $(2M+1)$ by 1 vector. Then, the first $(2M+2)$ moment conditions related to the equilibrium social interactions are found as

$$E[\frac{1}{N_r(N_r-1)}\sum_i\sum_j x_{ij,r}\hat{\nu}_{ij,r}] = 0, \tag{C.1}$$

$$E[\frac{1}{N_r(N_r-1)}\sum_i\sum_j d_{ij,r}\hat{\nu}_{ij,r}] = 0. \tag{C.2}$$

Note that the moment conditions are given for each network, and therefore we use the simple average of pairwise moments within a network. This average is not a sample analogue. We choose the simple average for simplicity. One can use a weighted average instead. The next $(2M+2)$ moment conditions are related to social capital.

$$E[\frac{1}{N_r}\sum_i\big(\sum_j x_{ij,r}\big)\hat{\xi}_{i,r}] = 0, \tag{C.3}$$

$$E[\frac{1}{N_r}\sum_i\big(\sum_j d_{ij,r}\big)\hat{\xi}_{i,r}] = 0. \tag{C.4}$$

Let $\bar{x}_i = \sum_j x_{ij}$ and $\bar{g}_R^{(m)}(\theta)$ be the mth sample moment:

$$
\begin{aligned}
\bar{g}_R^{(1)}(\theta) &= \frac{1}{R}\sum_r \Big[\frac{1}{N_r(N_r-1)}\sum_{j\neq i}\sum_i \hat{\nu}_{ij,r}(\theta)\Big], \\
\bar{g}_R^{(2)}(\theta) &= \frac{1}{R}\sum_r \Big[\frac{1}{N_r(N_r-1)}\sum_{j\neq i}\sum_i x_{ij,1,r}\hat{\nu}_{ij,r}(\theta)\Big], \\
&\vdots \\
\bar{g}_R^{(2M+1)}(\theta) &= \frac{1}{R}\sum_r \Big[\frac{1}{N_r(N_r-1)}\sum_{j\neq i}\sum_i x_{ij,2M,r}\hat{\nu}_{ij,r}(\theta)\Big], \\
\bar{g}_R^{(2M+2)}(\theta) &= \frac{1}{R}\sum_r \Big[\frac{1}{N_r(N_r-1)}\sum_{j\neq i}\sum_i d_{ij,r}\hat{\nu}_{ij,r}(\theta)\Big], \\
\bar{g}_R^{(2M+3)}(\theta) &= \frac{1}{R}\sum_r \Big[\frac{1}{N_r}\sum_i \hat{\xi}_{i,r}(\theta)\Big], \\
\bar{g}_R^{(2M+4)}(\theta) &= \frac{1}{R}\sum_r \Big[\frac{1}{N_r}\sum_i \bar{x}_{i,1,r}\hat{\xi}_{i,r}(\theta)\Big], \\
&\vdots \\
\bar{g}_R^{(4M+3)}(\theta) &= \frac{1}{R}\sum_r \Big[\frac{1}{N_r}\sum_i \bar{x}_{i,2M,r}\hat{\xi}_{i,r}(\theta)\Big], \\
\bar{g}_R^{(4M+4)}(\theta) &= \frac{1}{R}\sum_r \Big[\frac{1}{N_r}\sum_i \bar{d}_{i,r}\hat{\xi}_{i,r}(\theta)\Big].
\end{aligned}
$$

We collect all the sample moments in $\bar{g}_R(\theta)$. Then, the MSM estimator $\hat{\theta}_{MSM}$ minimizes the objective function

$$G(\theta) = \bar{g}_R(\theta)^{\mathrm{T}} W_R \bar{g}_R(\theta),$$

where W_R is a weighting matrix. We use the traditional two-step approach to estimate the weighting matrix, but given the large number of moments with a small sample size, we use a diagonal weighting matrix suggested by Pischke (1995). That is, we first use the identity matrix to obtain $\hat{W}_R$ and then run the optimization again by replacing the identity matrix with a weighting matrix with the optimal weights from $\hat{W}_R$ on the main diagonal and zeros elsewhere.

C.2 Many networks

The asymptotic argument in our MSM estimation is the number of networks R, which is often referred to the "many markets" asymptotic analysis in the empirical industrial organization literature. This approach has a few advantages. First, the many markets asymptotic analysis allows us to have a flexible correlation structure among $\varepsilon_{ij,r}$ within a network, which will be further discussed in the next subsection. Under the assumption of independent and identically distributed networks, the regularity conditions for MSM estimation in Pakes and Pollard (1989) are satisfied although pair-level observations in one network are not necessarily independent with each other. Therefore, the MSM estimator $\hat{\theta}$ is consistent and asymptotically normally distributed.

Second, the different levels of observations between the first and the second prediction errors do not create a problem. The population moment condition is constructed at the network level, and therefore, the pair-level and individual-level prediction errors are taken into account as the form of average. For example, the rth sample moment for the first moment condition for network r can be written as $g_r^{(1)}(\theta) = \frac{1}{N_r(N_r-1)} \sum_{j \neq i} \sum_i \hat{\nu}_{ij,r}(\theta)$.

The variance covariance matrix of the MSM estimator can be computed using either a formula given in Pakes and Pollard (1989) or a resampling method. We use the bootstrap method to compute standard errors and report them. We do so because the bootstrap standard errors are more efficient given that our sample size (the number of networks) is small ($R = 104$). The simulation errors remain fixed for all estimation procedures including bootstrap.

D Calibration in the policy exercises

Consider equations (44)–(48) 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

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

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

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

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: $\mathbf{\Delta}_r = (d_{ij,r})$ and $\mathbf{\Theta}_r = (\theta_{ij,r})$ as in (25). Thus, in vector-matrix form, (D.2) can be written as:

$$\mathbf{s}_r = \mathbf{1}_{N_r} + \frac{\alpha}{2b_r}\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)\mathbf{s}_r + \frac{\alpha\sigma_r N_r}{2b_r}\mathbf{1}_{N_r} - \frac{\alpha\,(1-\tau_r)\,c}{2b_r}\mathbf{\Delta}_r\mathbf{1}_{N_r}$$

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:

$$\mathbf{s}_r = \left[\mathbf{I}_{N_r} - \frac{\alpha}{2b_r}\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)\right]^{-1}\left[\left(1 + \frac{\alpha\sigma_r N_r}{2b_r}\right)\mathbf{1}_{N_r} - \frac{\alpha\,(1-\tau_r)\,c}{2b_r}\mathbf{\Delta}_r\mathbf{1}_{N_r}\right]$$

or equivalently

$$\mathbf{s}_r = \left[\mathbf{I}_{N_r} - \frac{\alpha}{2b_r}\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)\right]^{-1}\left[\left(1 + \frac{\alpha\sigma_r N_r}{2b_r}\right)\mathbf{I}_{N_r} - \frac{\alpha\,(1-\tau_r)\,c}{2b_r}\mathbf{\Delta}_r\right]\mathbf{1}_{N_r} \tag{D.3}$$

where $\mathbf{I}_{N_r}$ is the $(N_r \times N_r)$ identity matrix. The matrix $\mathbf{I}_{N_r} - \frac{\alpha}{2b_r}\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)$ is invertible if $\frac{\alpha}{2b_r} < \frac{1}{\rho(\mathbf{N}_0 + \mathbf{\Theta}_r)}$, where $\rho\left(\mathbf{N}_0 + \mathbf{\Theta}_r\right)$ is the largest eigenvalue of the matrix $\mathbf{N}_0 + \mathbf{\Theta}_r$. 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

Variable	Variable definition	(1) Mean (std.dev)	(2) Mean (std.dev)	Difference [P-value]	(3) Mean (std.dev)	Difference [P-value]	(4) Mean (std.dev)	Difference [P-value]
Female	Dummy variable taking value one if the respondent is female	0.51 (0.5)	0.5 (0.5)	[0.90]	0.53 (0.49)	[0.80]	0.58 (0.49)	[0.10]
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]	0.17 (0.4)	[0.15]	0.2 (0.4)	[0.37]
Other race	Dummy for other races	0.015 (0.12)	0.011 (0.1)	[0.42]	0.006 (0.08)	[0.65]	0.007 (0.08)	[0.37]
Student grade	Grade of student in the current year, range 7 to 12	9.67 (1.63)	9.49 (1.62)	[0.77]	9.48 (1.62)	[0.26]	9.18 (1.62)	[0.17]
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]	2.86 (0.71)	[0.07]	2.98 (0.71)	[0.22]
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]	2.37 (1.34)	[0.97]	2.12 (1.34)	[0.42]
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 (1.09)	[0.79]	3.27 (1.09)	[0.93]
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]	0.72 (0.45)	[0.42]	0.72 (0.45)	[0.93]
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 (0.9)	[0.46]	3.19 (0.9)	[0.45]
Family income	Family income in thousands of dollars	40.72 (50.76)	39.93 (50.32)	[0.16]	44.38 (73.09)	[0.10]	44.56 (73.09)	[0.34]
Family size	Number of people living in the household	3.61 (1.67)	3.52 (1.51)	[0.80]	3.39 (1.33)	[0.07]	3.32 (1.33)	[0.65]
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]	0.1 (0.31)	[0.30]	0.11 (0.31)	[0.67]
N.obs		20,745	12,761		2,236		890	

Note: (1): original sample, (2): sample with geocoded information, (3): Sample with social-interaction information (network size 2–460), (4) Sample in networks of size 4–70. 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)	(2)	(3)	(4)
**** Social interaction equation ****				
β**: social distances**	$(x_i - x_j)$	$(x_i - x_j)$	$\lvert x_i - x_j \rvert$	$\lvert x_i - x_j \rvert$
Female	-0.00040**	-0.00033**	-0.00024**	-0.00022**
	(0.00018)	(0.00014)	(0.0001)	(0.00009)
Black	-0.0011**	-0.0019**	-0.0147**	-0.0127**
	(0.0005)	(0.0008)	(0.0072)	(0.0061)
Grade	-0.0040**	-0.0035**	-0.0241**	-0.0233**
	(0.0017)	(0.0014)	(0.0102)	(0.0106)
GPA	-0.0046**	-0.0037**	-0.0017**	-0.0017**
	(0.0023)	(0.0015)	(0.0007)	(0.0007)
Parental education	0.0069**	0.0077**	0.0035**	0.0034**
	(0.0031)	(0.0032)	(0.0016)	(0.0013)
Family income	-0.0046**	-0.0020**	0.0013**	0.0012**
	(0.0021)	(0.0008)	(0.0005)	(0.0005)
Two parents	0.00044**	0.00034**	-0.00335**	-0.00327*
	(0.00022)	(0.00014)	(0.00156)	(0.00177)
Other race		-0.0145**		-0.0021***
		(0.0058)		(0.0008)
Physical development		0.0041**		-0.0031**
		(0.0017)		(0.0013)
Religion practice		-0.0039**		-0.0189**
		(0.0016)		(0.0086)
Family size		0.0040**		-0.0011**
		(0.0016)		(0.0005)
Family income refused		0.0031**		-0.0057**
		(0.0013)		(0.0024)
β**: combined levels**	$(x_i + x_j)$	$(x_i + x_j)$	$(x_i + x_j)$	$(x_i + x_j)$
Female	0.0031**	0.0048**	-0.0028**	-0.0030**
	(0.0012)	(0.002)	(0.0013)	(0.0013)
Black	0.0196**	0.0206**	0.0066**	0.0062**
	(0.0095)	(0.0087)	(0.0032)	(0.0025)
Grade	-0.00037***	0.00642***	-0.00813***	0.0377**
	(0.00012)	(0.00247)	(0.00237)	(0.0147)
GPA	-0.0012***	-0.0117**	-0.0212***	-0.0291**
	(0.0004)	(0.0047)	(0.0065)	(0.0132)
Parental education	-0.00042***	0.00313**	0.00082***	0.00092**
	(0.00015)	(0.00124)	(0.00024)	(0.00039)
Family income	-0.0049**	-0.0037**	-0.0047**	-0.0047**
	(0.0023)	(0.0015)	(0.0023)	(0.0020)
Two parents	0.0037**	0.0053**	0.00009**	0.00008**
	(0.0015)	(0.0022)	(0.00004)	(0.00004)
Other race		0.0062**		0.0097***
		(0.0026)		(0.0033)
Physical development		0.0043**		0.0017**
		(0.0017)		(0.0008)
Religion practice		0.0054**		-0.0051**
		(0.0022)		(0.0021)
Family size		-0.0048**		-0.0097**
		(0.0019)		(0.0040)
Family income refused		0.0066**		-0.0014**
		(0.0027)		(0.0005)
n_0	2.7713**	2.4992**	3.0630***	2.2698***
	(1.1191)	(1.0166)	(1.1723)	(0.7849)
c (transportation cost)	0.00014**	0.00011**	0.00026*	0.00023**
	(0.00007)	(0.00005)	(0.00014)	(0.0001)
σ_ϵ	3.9449**	3.5152**	4.5948**	2.667**
	(1.7992)	(1.4374)	(1.8119)	(1.0471)
**** Social capital equation ****				
α	0.0334**	0.0243**	0.0167*	0.0179**
	(0.0134)	(0.0100)	(0.0091)	(0.0076)
Number of networks	104	104	104	104
Number of pupils	890	890	890	890
Number of directed pairs	18,482	18,482	18,482	18,482

Note: We estimate parameters $(n_0, c, \beta^{\mathrm{T}}, \sigma_\varepsilon)^{\mathrm{T}}$ in the social interaction equation (19) and subsequent specifications (22)–(21), and parameter α in the social capital equation (20).
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 4: Social interactions and geographic dispersion of students

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

Note: Dispersion of a network is measured by taking the average of distances from each student's home to the network center.
Average distance is the average of pairwise distances of students in a network.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5: 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
2.349	1.144	1.205	-0.810	2.741	1.265	1.028	0.236	-0.095	3.504

Note: The statistics are computed using the network-level average social interactions and social capital from 104 networks. For example, the largest difference between the average levels of optimal and observed social interactions is 2.741. Note that these statistics differ from pair-level averages.
The observed level of social capital is augmented using equation (27).

Table 6: Difference between optimal level and observed level of social interactions

	Optimal−Observed (social interactions)				
	(1)	(2)	(3)	(4)	(5)
Network population	0.1060***	0.1020***	0.143***	0.0637***	0.0614***
	(0.0109)	(0.0109)	(0.0149)	(0.0190)	(0.0130)
Network population^2	-0.0016***	-0.0015***	-0.0019***	-0.0011***	-0.0011***
	(0.00018)	(0.00018)	(0.00021)	(0.00021)	(0.00016)
Avg. geographic distance		0.0145*	-0.0078	-0.0130**	-0.0094**
		(0.0073)	(0.0060)	(0.0060)	(0.0041)
Avg. degree centrality			-0.260***	-0.107**	-0.171***
			(0.0334)	(0.0524)	(0.0357)
Std.dev. of degree centrality			-0.0178	-0.1130**	-0.0397
			(0.0393)	(0.0442)	(0.0285)
Avg. eigenvector centrality				-5.156***	-4.477***
				(1.022)	(0.639)
Clustering coefficient				-1.080	0.583
				(1.162)	(0.646)
Diameter				-0.0394**	-0.0243**
				(0.0158)	(0.0107)
Female fraction					-0.0708
					(0.0829)
Black fraction					0.0037
					(0.0824)
Other race fraction					0.7810
					(0.6020)
Avg. student grade					0.0724***
					(0.0228)
Avg. GPA					-0.1150*
					(0.0682)
Avg. level of religion practice					-0.0281
					(0.0294)
Avg. family size					-0.0888**
					(0.0363)
Fraction of students with two parents					0.0703
					(0.1070)
Avg. level of physical development					0.0054
					(0.0493)
Avg. family income					-0.0060***
					(0.00057)
Fraction family income refused					0.2290
					(0.1600)
Constant	0.583***	0.487***	1.774***	4.263***	4.103***
	(0.091)	(0.110)	(0.144)	(0.478)	(0.396)
Observations	104	104	104	104	104
R-squared	0.357	0.374	0.618	0.687	0.915

Note: The outcome variable is the difference between optimal level and observed level of social interactions for each network. Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 7: Difference between optimal level and observed level of social capital

	Optimal−Observed (social capital)				
	(1)	(2)	(3)	(4)	(5)
Network population	0.0115	0.0177*	0.0157	-0.0120	-0.0129
	(0.0087)	(0.0099)	(0.0111)	(0.0232)	(0.0246)
Network population^2	-0.000215	-0.000314**	-0.000294**	0.000019	0.000047
	(0.00013)	(0.00015)	(0.00015)	(0.00025)	(0.00027)
Avg. geographic distance		-0.0217**	-0.0211**	-0.0208**	-0.0194*
		(0.0103)	(0.0105)	(0.0102)	(0.0108)
Avg. degree centrality			0.00022	-0.0118	-0.0139
			(0.0414)	(0.0514)	(0.0593)
Std.dev. of degree centrality			0.0087	0.0175	0.0280
			(0.0688)	(0.0660)	(0.0638)
Avg. eigenvector centrality				-1.011	-0.704
				(0.838)	(1.018)
Clustering coefficient				-0.929	-0.625
				(0.910)	(0.949)
Diameter				0.0124	0.0149
				(0.0199)	(0.0217)
Female fraction					-0.0743
					(0.152)
Black fraction					0.0517
					(0.108)
Other race fraction					0.165
					(1.268)
Avg. student grade					-0.0404
					(0.0494)
Avg. GPA					-0.120
					(0.178)
Avg. level of religion practice					0.0553
					(0.0670)
Avg. family size					-0.0409
					(0.0538)
Fraction of students with two parents					0.248
					(0.184)
Avg. level of physical development					0.0473
					(0.136)
Avg. family income					-0.00094
					(0.00087)
Fraction family income refused					0.254
					(0.323)
Constant	0.184**	0.328**	0.303*	0.764*	1.047
	(0.0760)	(0.134)	(0.162)	(0.436)	(0.753)
Observations	104	104	104	104	104
R-squared	0.007	0.071	0.071	0.088	0.117

Note: The outcome variable is the difference between optimal level and observed level of social capital for each network. The observed level of social capital is augmented using equation (27).
Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table 8: Optimal network design: average welfare and number of students

	(1) Welfare	(2) Welfare	(3) Welfare	(4) Welfare	(5) Welfare
Network population	-2.108*	-1.074	0.488	3.590	2.071
	(1.245)	(0.918)	(1.514)	(3.762)	(3.264)
Network population^2	0.0279	0.0114	-0.0040	-0.0402	-0.0187
	(0.0178)	(0.0133)	(0.0150)	(0.0394)	(0.0327)
Avg. geographic distance		-3.610**	-3.914*	-3.968*	-3.181*
		(1.755)	(2.117)	(2.148)	(1.834)
Avg. degree centrality			1.835	4.519	3.679
			(3.299)	(4.774)	(9.706)
Std.dev. of degree centrality			-8.068	-9.581	-4.249
			(9.719)	(9.643)	(10.05)
Avg. eigenvector centrality				87.22	158.5
				(116.4)	(194.9)
Clustering coefficient				3.164	76.42
				(97.79)	(198.0)
Diameter				-1.734	-0.805
				(1.521)	(2.489)
Controls	No	No	No	No	Yes
Observations	104	104	104	104	104
R-squared	0.005	0.025	0.051	0.060	0.132

Note: Control variables include the averages of the social distances and the combined levels used in structural estimation. See Table 2.
Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9: Policy levels for optimal outcomes

(1) Subsidizing social interactions: σ			(2) Subsidizing transportation costs: τ		
Average	Minimum	Maximum	Average	Minimum	Maximum
0.4133	-0.5473	10.3352	0.9471	-0.2470	17.7655

Note: The subsidy level for each network is computed for students in each network to obtain the optimal level of social interactions and social capital in (36)–(38).

Table 10: Comparison of two policies

	Number of networks that lead to higher welfare for each policy	
Subsidy schemes	Policy: σ	Policy: τ
(1) Uniform subsidy amount for each network	81	11
(2) Subsidy proportional to N_r	81	11
(3) Subsidy proportional to $N_r(N_r-1)$	97	7

Note: 104 networks. Two policies are tied for some networks.