Community Size and Network Closure

By Hunt Allcott, Dean Karlan, Markus M. Möbius, Tanya S. Rosenblat, and Adam Szeidl*

Large community size is often negatively correlated with prosocial behaviors such as formal volunteering, working on public projects, and informal help to friends and strangers (Robert D. Putnam 2000, 119, 206). This may be because people who reside in large communities simply spend less time socializing with each other. As a result, people living in large cities have on average fewer friends, and their social networks support less cooperation.[1]

A complementary channel, which has received less attention in the literature, is that community size may affect outcomes by changing other aspects of the network structure. Specifically, even holding fixed the number of friends, we expect social networks in small communities to exhibit greater *network closure*, i.e., be more interconnected. The intuition is straightforward. In small communities, the pool of potential friends is limited, which increases the extent to which the network neighborhoods of two friends are likely to overlap.

James S. Coleman (1990) suggested that variation in network closure can affect outcomes. In particular, he argued that networks with higher closure generate high trust between friends, which facilitates cooperation and improves welfare. The logic is that networks with high closure allow for greater social sanctions between individuals through common friends, increasing incentives for cooperation.

In this paper, we empirically evaluate the effect of community size on network closure in one specific social environment—schools. We make use of the National Longitudinal Study of Adolescent Health (AddHealth), which has detailed social network information for about 86,000 students in 142 US middle and high schools. Our main finding is a strong negative relationship between grade size and network closure. We also find that grade size is associated negatively with prosocial behavior.[2] We then explore whether this community size effect on outcomes can be partially attributed to the network closure channel.

I. Theory and Estimation Framework

A. *Community Size and Network Structure*

Community size can affect network structure by influencing the evolution of friendship links between agents. To see the logic, imagine that the social network is generated by a random process where: (a) agents' preferences for socializing with other agents are i.i.d. draws from a known distribution, and (b) agents are more likely to interact with those who are around them (see, e.g., Matthew O. Jackson and Brian W. Rogers 2006). With such a process, the degree distribution measuring the number of friends per agent will be independent of community size. The friends of an agent in a small community will likely be friends with each other, however, because there are few other potential friends. This generates

* Allcott: Harvard University, 1050 Massachusetts Avenue, Cambridge, MA 02138, and National Bureau of Economic Research (e-mail: hunt_allcott@ksg.harvard.edu); Karlan: Department of Economics, Yale University, 77 Prospect Street, New Haven, CT 06520 (e-mail: dean.karlan@yale.edu); Möbius: Department of Economics, Harvard University, Littauer Center, Cambridge, MA 02138, and NBER (e-mail: mobius@fas.harvard.edu); Rosenblat: Department of Economics, Wesleyan University, PAC 123, 238 Church Street, Middletown, CT 06459 (e-mail: trosenblat@wesleyan.edu); Szeidl: Department of Economics, University of California, Berkeley, Berkeley, CA 94720 (e-mail: szeidl@econ.berkeley.edu). We thank Avinash Dixit for comments.

[1] For example, Jane Jacobs (1961) criticized city planning of the 1950s and 1960s, with its emphasis on large blocks, because it prevented many types of social interactions common in small cities, such as convening on sidewalks and in small squares.

[2] In the education literature, there is substantial agreement that smaller schools are associated with better outcomes (Kathleen Cotton 2001). Possible mechanisms beyond the closure channel include teacher attitudes (Valerie E. Lee and Susanna Loeb 2000), improved safety (Patricia A. Wasley et al. 2000), decision-making autonomy (Wasley et al. 2000), and more parent involvement (Stacy Mitchell 2000). In economics, Caroline M. Hoxby (2000) and Joshua D. Angrist and Victor Lavy (1999), among others, show the connection between *class size* and student achievement.

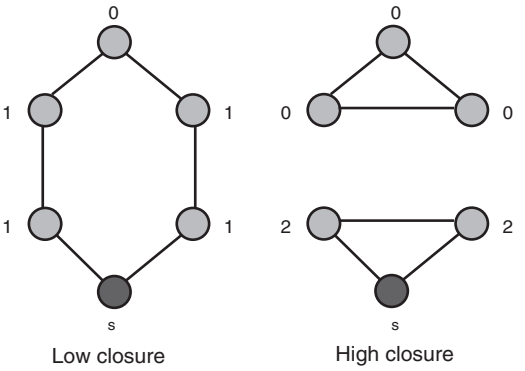

FIGURE 1. LOW-CLOSURE (LEFT) AND HIGH CLOSURE (RIGHT) SOCIAL NETWORKS WITH IDENTICAL LINK DISTRIBUTION (K = 1.5)

Notes: Agent *s* enjoys the same amount of average trust in both networks. In the low-closure network she can ask four agents for favors worth at most one, while in the high-closure network she can ask two agents for favors worth at most two.

overlapping neighborhoods and high network closure. In contrast, in a large community the friends of an agent are likely to have different friends, leading to nonoverlapping neighborhoods and low closure. This model predicts that the social networks in small communities should exhibit higher network closure.

B. *A Measure of Network Closure*

To assess empirically the magnitude of the channel described above, we will use the measure of network closure developed by Möbius and Szeidl (2006). We now explain and provide intuition for this measure.

First, we introduce the idea of a trust flow, which measures the degree to which agents in a social network can cooperate. Define the trust flow between two agents *s* and *t* to be the number of disjoint paths in the network connecting them, where all edges in a path must be within distance K from agent *s*.[3] Here *K* governs the size of the "circle of trust," i.e., the set of agents who can be used to enforce cooperation. Intuitively, high trust flow means that *s* and *t* have better incentives to cooperate because they have access to greater social sanctions over each other should one of them defect.

To measure network closure for a given circle of trust *K*, consider an agent *s*, and compute the trust flow between *s* and all other agents in the network. Let the sum of these pairwise flows be the total trust of agent *s*, which measures how much cooperation *s* can expect from the network. For any positive integer *m*, compute the share of this total trust that comes from pairwise trust flows exceeding *m*. The resulting number, which we refer to as $(m,\ K)$-closure of agent *s*, is a measure of the closure of the network around agent *s*. Loosely speaking, high $(m,\ K)$-closure can be interpreted as *s* having high pairwise trust with a small number of agents, while lower closure means having low pairwise trust with a large number of agents.

It is important to note that closure can vary across networks even if the network size and degree distribution are held fixed. For example, Figure 1 shows two networks where each of the six agents has two friends. Agent *s* has lower closure in the network on the left, however. To see this formally, let $K = 1.5$. The trust flows between *s* and all other agents for this *K* are shown in the figure. In both networks, *s* enjoys a total trust of 4. In the left network this total trust comes from low pairwise trust flows of one from four other agents, while in the right network it comes from high pairwise trust flows of two from two other agents. Hence, agent *s* has $(1, K)$-closure of zero in the left network, and $(1, K)$-closure of one in the right network.

Is high or low closure better for agent *s*? That depends on what he wants to use the network for. If he needs the network for valuable transactions, then high closure is better. While high closure provides access to only a few agents, *s* can ask valuable favors from all of these. In contrast, if *s* needs to engage in transactions of lower value, he prefers the greater access to people provided by the low-closure network.

C. *Estimation Framework*

To evaluate the effect of community size on network closure, we will estimate equations of the form

$$\text{(1)}\quad \text{closure} = \alpha + \beta \cdot \text{community size} + \gamma \cdot \text{controls} + \varepsilon,$$

[3] The distance between an agent and an edge is the average of the distances between the agent and the two end points of the edge.

where β measures the causal effect of community size on network closure, ε captures unobserved variation and measurement error, and the controls refer to other social and economic factors that can influence network structure. One important control variable is an agent's number of friends, also called degree. Degree matters because it can be related to both community size and closure. More specifically, community size may affect the network structure through influencing degree. For example, if people in small communities invest more in social interactions, they end up with more friends. Alternatively, people in large communities may find a bigger potential market for friends, and, as a result, develop more connections. This effect of community size on degree can bias the estimation of (1) because degree is correlated positively with closure. Networks with higher average degree tend to be more interconnected: to deal with this difficulty, we include flexible controls for degree in our specifications.

We will also be interested in the effects of the variation in closure generated by community size on social and educational outcomes. To assess the importance of this channel, we first estimate the total effect of community size on outcomes using the following specification:

$$\text{(2)} \quad \text{outcome} = \delta + \theta \cdot \text{community size} + \gamma \cdot \text{controls} + \nu.$$

Here, the total effect of size is measured by θ. To assess how much of this effect is due to variation in network closure, we then estimate

$$\text{(3)} \quad \text{outcome} = \delta' + \theta' \cdot \text{community size} + \rho \cdot \text{closure} + \gamma' \cdot \text{controls} + \upsilon' + (\nu').$$

If part of the effect of size on outcomes is mediated through network closure, then we expect that the absolute value of the coefficient on size falls as we include closure in the regression ($|\theta'| < |\theta|$). Moreover, using the estimated coefficients in these regressions, we can measure the relative contribution of the network closure channel in the total effect of size on outcomes.

This analysis, like others using school size as an independent variable, is frustrated by the lack of exogenous variation in grade size. Our estimation procedure requires that the error terms in regressions (1), (2), and (3) are orthogonal to the key size and closure variables. One possible concern with this assumption is that in equation (1), size may be correlated with closure beyond its causal effect, for example, if size is related to other environmental factors that affect network structure. Similarly, in equation (3), closure may be correlated with the outcome variable through an omitted variable, like the sociability of the school's teachers. We attempt to alleviate these concerns by including a rich set of controls, such as social and economic factors, in all regressions. We cannot fully rule out alternative explanations, however, and hence we interpret the findings from the analysis about closure and outcomes above as suggestive evidence.

II. Data Description

AddHealth surveyed every student in a representative sample of 142 US public and private middle and high schools in 1994 and 1995 (Richard J. Udry 2003). Each of these 85,627 students reported demographics, recent grades and extracurricular activities, health outcomes, and their feelings about their school environment.

As part of the survey, students were asked to name up to ten friends (five male and five female). We use this information to construct an "OR-social network," where a direct link between two students exists if one of them names the other one as a friend. We obtain our core sample by excluding 7,779 students who are in connected components of size smaller than 20 (7,138 form clusters of size 1). This is because network statistics become less meaningful for such small components. In particular, closure is undefined for clusters of size 1.

We compute our trust flow measure for $K = 2.0$ for all pairs of agents in the core sample, as well as the (m, K)-closure measure for $m = 2$ developed above.[4] We also compute network degree—the number of neighbors for each agent.

We measure community size by the number of students in the student's grade at her school.[5]

[4] Using $m = 1$ does not affect our qualitative results, nor does using $K = 2.5$.

[5] In most American high schools, students attend multiple classes with different groups of students, so we expect

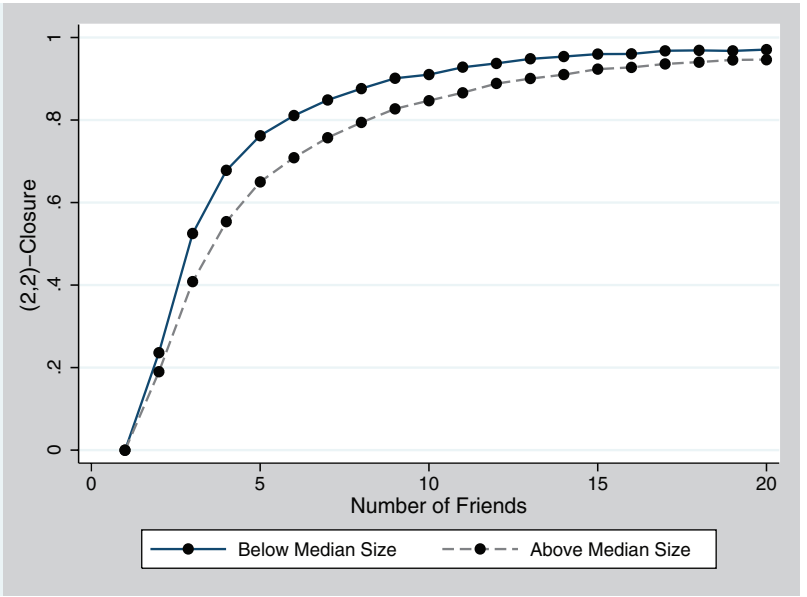

Figure 2. Mean (2, 2)-Closure by Number of Friends for Schools with Grade Size below Median Size (solid line) and Schools with Grade Size above Median Size (dashed)

Notes: See the text for definition of (2, 2)-closure.

We use categorical indicators to control for four observables: the student's grade (6 to 12), her ethnicity (white or nonwhite), her father's education (less than high school, high school, college) and her school's location (urban, suburban, or rural). We also include controls for fractionalization in some regressions. For grade, ethnicity, and education, fractionalization is the probability that two people chosen randomly from the cluster will be of different categories.

We focus on three outcome variables. TROUBLE is the student's response to the question, "Since school started this year, how often have you had trouble getting along with other students?," with 0 being "never" and 4 being "every day." FEELSAFE is the student's response to the question, "How strongly do you agree or disagree with the following statement: 'I feel safe in my school'," which we have coded from 1 ("strongly disagree") to 5 ("strongly agree"). GPA is the average of the student's most recent grades in English, math, history, and science.

their social networks to be influenced by school size rather than by classroom size. Our focus on school size distinguishes the current paper from the body of econometric work on class size.

III. Results

A. *Community Size and Network Closure*

We begin with a graphical analysis. Figure 2 plots our measure of network closure for schools with grade size above and below the median size, holding fixed a student's number of friends (degree). The figure provides preliminary evidence for our main hypothesis by showing that in our data, large schools tend to have lower network closure. It is important to note that this correlation between size and closure holds even when we control for number of friends. While degree and closure are correlated with each other, grade size generates variation in closure beyond its effect on the degree distribution.

These observations are also reflected in the regression analysis in Table 1, where we estimate (1) in different specifications. Increasing the number of students in a grade is correlated

TABLE 1—OLS AND FIXED EFFECTS REGRESSIONS OF NETWORK (2, 2)-CLOSURE ON GRADE SIZE

	OLS	FE	FE-controls	Fract-controls
	(1)	(2)	(3)	(4)
Student in grade	-0.040	-0.032	-0.026	-0.026
	(0.0007)**	(0.0004)**	(0.0005)**	(0.0005)**
Const.	0.822	0.800	0.770	0.722
	(0.002)**	(0.001)**	(0.002)**	(0.011)**
Obs.	77,848	77,848	77,848	77,848
R^2	0.036	0.069	0.095	0.095

Notes: Standard errors are in parentheses. See the text for definition of (2, 2)-closure. Fixed effect is for number of friends of student. Unit of grade size is 100 students. In column 3, we control for student's grade, school location (rural, suburban, and urban), parents' education, and racial composition. In column 4, we additionally include the demographic fractionalization controls.

**Significantly different from 0 at the 1 percent level.

TABLE 2—EFFECT OF GRADE SIZE ON OUTCOMES FOR FEELSAFE (FS), TROUBLE (TR), AND GPA, WITHOUT AND WITH (2, 2)-CLOSURE

	FS1	FS2	TR1	TR2	GPA1	GPA2
	(1)	(2)	(3)	(4)	(5)	(6)
(2,2)-closure		0.315		−0.415		0.427
		(0.025)**		(0.033)**		(0.018)**
Students in grade	−0.029	−0.021	−0.021	−0.032	−0.033	−0.022
	(0.003)**	(0.003)**	(0.004)**	(0.004)**	(0.002)**	(0.002)**
Const.	3.747	3.512	1.697	2.005	2.688	2.368
	(0.075)**	(0.078)**	(0.098)**	(0.101)**	(0.056)**	(0.058)**
Obs.	69,042	69,042	73,780	73,780	68,220	68,220
R^2	0.022	0.024	0.04	0.042	0.047	0.054

Notes: Standard errors are in parentheses. All columns include fixed effects for number of friends. Unit of grade size is 100 students. All regressions include controls for student's grade, school location (rural, suburban, and urban), parents' education, racial composition, as well as demographic fractionalization.

**Significantly different from 0 at the 1 percent level.

with lower closure both in the simple ordinary least squares (OLS) regressions and in specifications where we include fixed effects for the number of friends of a student. Including a rich set of controls such as grade composition within the connected component, school location, parents' education, racial composition, and demographic fractionalization does not change our results. The fixed effect estimates suggest that a one-standard-deviation increase in grade size is associated with a decrease in closure by about 0.12 of a standard deviation.

B. *Closure and Outcomes*

Table 2 presents estimates of equations (2) and (3) for our three outcome variables (TROUBLE, FEELSAFE, and GPA). In all specifications we include fixed effects for number of friends, as well as the full set of controls used in the fourth column of Table 1.

Columns FS1 and GPA1 show that students in larger schools feel less safe and also have significantly lower GPA. These results are consistent with the notion that grade size is negatively related to educational and social outcomes. Somewhat surprising, however, is the negative relationship between having trouble with other students and grade size (column TR1). A one-standard-deviation increase in grade size is associated with about a 0.04-standard-deviation decrease in the FEELSAFE measure, and a 0.06-standard-deviation decrease for GPA.

To assess the extent to which the effect of grade size on outcomes is mediated through network closure, we include closure in the right-

hand side of our regressions. As columns FS2, TR2, and GPA2 show, closure is positively correlated with good outcomes. The coefficient of closure is strongly significant and has the right sign in all three specifications. Moreover, focusing on the FEELSAFE and GPA outcomes where grade size predicts worse outcomes, we also find that including closure significantly reduces the grade size coefficient. The drop in the effect of grade size is between one-fourth and one-third for these two outcomes. These results suggest that a substantial part of the effect of size on FEELSAFE and GPA is mediated through the closure channel.

We interpret our outcome estimates as suggestive evidence supporting Coleman's (1990) claim that higher closure generates prosocial behavior. Mobius and Szeidl (2006) predict such a positive relationship if agents engage in high-value transactions within the network. Coming to the aid of fellow students or helping friends who lag behind academically might qualify as such "high-value" favors.

IV. Conclusion

In this paper, we argue that community size can affect outcomes by influencing the interconnectedness of social networks. Using the AddHealth dataset, we show that there is a strong association between grade size and network closure, which is robust to controlling for the degree distribution of the social network. We also provided suggestive evidence that this relationship between size and closure can influence trust-related outcomes.

Unfortunately, the AddHealth dataset contains a small number of outcomes where the role of the social network is likely to be important. It would be interesting to apply our methodology to data with more outcome variables. In particular, we propose dividing the analysis by "high-value" and "low-value" services provided by the social network (such as financial support versus information exchanged with a social network). This would allow us to test whether networks with lower closure can sometimes provide greater utility to agents who have a demand for "low-value" services (as suggested by Ronald S. Burt 1995).

REFERENCES

Angrist, Joshua D., and Victor Lavy. 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement." *Quarterly Journal of Economics*, 114(2): 533–75.

Burt, Ronald S. 1995. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Cotton, Kathleen. 2001. "New Small Learning Communities: Findings from Recent Literature." Unpublished.

Hoxby, Caroline M. 2000. "The Effects of Class Size on Student Achievement: New Evidence from Population Variation." *Quarterly Journal of Economics*, 115(4): 1239–85.

Jackson, Matthew O., and Brian W. Rogers. Forthcoming. "Meeting Strangers and Friends of Friends: How Random are Socially Generated Networks?" *American Economic Review.*

Jacobs, Jane. 1961. *The Death and Life of Great American Cities*. New York: Random House.

Lee, Valerie E., and Susanna Loeb. 2000. "School Size in Chicago Elementary Schools: Effects on Teachers' Attitudes and Students' Achievement." *American Educational Research Journal,* 37(1): 3–31.

Mitchell, Stacy. 2000. "Jack and the Giant School." *The New Rules*, 2: 1–10.

Mobius, Markus M., and Adam Szeidl. 2006. "Trust and Social Collateral." Unpublished.

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

Udry, Richard J. 2003. "The National Longitudinal Study of Adolescent Health (AddHealth), Waves I and II 1994–1996; Wave III, 2001–2002." Carolina Population Center, University of North Carolina at Chapel Hill. http://www.cpc.unc.edu/projects/addhealth/data.

Wasley, Patricia A., Michelle Fine, Matt Gladden, Nicole E. Holland, Sherry P. King, Esther Mosak, and Linda C. Powell. 2000. *Small Schools: Great Strides. A Study of New Small Schools in Chicago.* New York: The Bank Street College of Education.