What Makes a Classmate a Peer?
Examining which peers matter in NYC elementary schools

William C. Horrace, Hyunseok Jung, Jonathan L. Presler, Amy Ellen Schwartz *

August 2022

Abstract

This paper identifies and estimates the effects of student-level social spillovers on standardized test performance in New York City (NYC) elementary schools. We leverage student demographic data to construct within-classroom social networks based on shared student characteristics, such as a gender or ethnicity. Rather than aggregate shared characteristics into a single network matrix, we specify additively separate network matrices for each shared characteristic and estimate city-wide peer effects for each one. Conditional on being in the same classroom, we find that the most important student peer effects are shared ethnicity, gender, and primary language spoken at home. We show that altering classroom composition changes the impact of these networks. Particularly, low ethnic diversity is correlated with low impact for shared ethnicity. We discuss identification of the model and its implications for within- and between-group test performance gaps along several demographic traits.

Keywords: Peer effect, Network, Homophily, Education.
JEL Codes: C31, I21

NELSON MUNTZ: *Ha ha! Your mom's a jailbird!*
BART SIMPSON: *So is yours.*
NELSON MUNTZ: *Oh, yeah. Let's play!*
(The Simpsons, Season 4, Episode 21)

Acknowledgements

We gratefully acknowledge funding from the Institute of Education Sciences (award number: R305A170270). We would like to thank the NYC Department of Education for providing data and for their support, especially the Office of Pupil Transportation, Alexandra Robinson, and Tim Calabrese. For data support, advice, and suggestions, we also thank Meryle Weinstein, Sarah Cordes, and Joanna Bailey. We thank Stephen Ross, seminar participants at the Maxwell School at Syracuse University, the AEFP conference, and Daniel Patrick Moynihan Summer Workshop in Education and Social Policy for useful comments on previous drafts. The opinions expressed are those of the authors and do not represent views of the NYC Department of Education.

*Corresponding author: Hyunseok Jung, hj020@uark.edu. Horrace: Syracuse University. Jung: University of Arkansas. Presler: Saint Louis University. Schwartz: University of Delaware

1 Introduction

Birds of a feather flock together. This concept is called *homophily*, and we see its effects in schools, work, and social circles - both personally and in popular culture. The effects of homophily can be quite stark. For example, compared to their male counterparts, female investors are three times as likely to invest in companies with a female CEO.[1] Homophily may arise if external similarities correlate with shared cultural experiences, making potential relationships less costly through improved information flows and efficient communication.[2] On the other hand, homophily could simply be the result of taste, which may correlate with prejudice or fear of the other.[3] This paper asks two questions related to homophily in elementary school classroom in New York City. First, which demographic characteristics (if any) are important for social spillovers in standardized test performance, and how do they compare in sign and magnitude? For example, is shared gender more influential than shared neighborhood of residence? Second, what are the effects of such homophily-induced spillovers on the distribution of student academic performance within and between demographic groups? Specifically, how do these spillovers affect existing achievement gaps within the classroom?

Understanding how homophily works, and which networks are important in which contexts, is important for education policymakers wishing to harness peer effects (or, at least, to better understand them). The promise of peer effects is twofold: a near costless improvement to educational outcomes through optimal assignment of students to classrooms and a "social multiplier effect" for exogenous policy interventions and investments.[4] A major challenge to policy intervention is that students sort into groups within the classroom, and this sorting behavior may bring unintended consequences when combined with an intervention.[5] This

[1] Abramson et al. (2019) discusses gender diversity in venture capital, and its implications.

[2] See Hegde and Tumlinson (2014) for a model of this in venture capital firms. They discuss how ethnic homophily plays into both selection of which companies to invest in, as well as influence after investment.

[3] See Leszczensky and Pink (2019) for a discussion of how different taste preferences for homophily (high and low identifiers) interact to form groups of varying levels of homogeneity.

[4] Bennett and Bergman (2018) provide a nice example of social multipliers in action using an attendance intervention. A good survey of the related theory can be found in Epple and Romano (2011).

[5] See Carrell et al. (2013) for an example. The authors conducted an experiment in which they manipulated peer groups such that it was expected to help the lowest performing students. The target group of students

paper sheds light on this complication, by identifying and empirically testing for characteristics which are important for within-classroom spillovers and then by exploring how these effects change under differing classroom compositions.

To explore these effects, we use student-level administrative data from NYC, where we partition the universe of elementary school students into classrooms. Within each classroom, we consider seven *demographic partitions* (gender, ethnicity, neighborhood, bus stop, bus route, language spoken at home, and country of birth.)[6] Within each *demographic partition*, students are divided into *demographic groups* corresponding to the possible categories within each partition. For example, in any classroom there may be two gender groups (boys and girls) and four ethnicity groups (Black, White, Hispanic, and Asian/Other), and each student belongs to one of the two gender groups and one of the four ethnicity groups. We leverage variability in the size and composition of demographic groups in each classroom to identify peer effects through each demographic partition. That is, we simultaneous estimate a gender peer effect, an ethnicity peer effect, a neighborhood peer effect, etc. for elementary school students in NYC. A number of papers have studied homophily, but typically only a few demographic characteristics (e.g., gender and ethnicity) are considered *individually*. We later partition the data based on ethnic diversity in the classroom and re-estimate our model in order to explore how classroom contribution changes the relative effects. Is the ethnicity network more or less important in an ethnically diverse classroom than a nearly homogeneous classroom? Does the influence of other networks which may correlate with ethnicity vary with classroom diversity?

The main source of identification for our model is the variation in partition structures across demographic partitions.[7] In our model, groups of one partition *partially* overlap groups of the others when each partition uniquely divides students. This creates partially

ended up being hurt by the experiment because the way peer groups formed changed when the set of potential peers changed.

[6]For clarity of exposition, we refer to all partitions as *demographic partitions*, despite a few of them being based on geography.

[7]A partition structure indicates the way a demographic partition divides students in a classroom. Variation in partition structures exist when partitions are unique, or only partially overlapping.

or indirectly connected groups of students,[8] which generates exclusion restrictions similar to those by "transitive triads" (Bramoullé et al., 2009) and "partially overlapping groups" (Laschever, 2013; De Giorgi et al., 2010) in the network literature. In addition to these exclusion restrictions, variation in group sizes also helps identification of our model. Lee (2007) considers classroom interaction and shows that identification is possible based on classroom size variation. The same identification mechanism works for our model, but our model partitions classmates further into multiple demographic groups, which creates much richer variation in group sizes. We discuss how these variations yield identification using simple examples and existing identification results from the literature. In doing so, we address identification challenges common to the peer effects literature, such as reflection, selection, and correlated effects.

Empirically, our model finds the strongest networks are shared ethnicity, gender, and primary language spoken at home. Ethnicity groups are most important for mathematics test scores, and gender groups are most important for reading test scores. To our knowledge, this is the first paper to explore bus spillovers in the classroom, but we do not find them to be an important factor in classroom performance. In general, peer effects based on homophily appear to be stronger in mathematics than reading test scores. We explore different classroom compositions to understand how network influence of demographic partitions change when peer group composition varies. We provide evidence that low ethnic diversity leads to lower relevance for the ethnicity network.

In our model, group interaction along multiple demographic partitions produces very general and interesting reduced-form dynamics compared to existing models. First, positive within-group interaction may decrease existing within-classroom performance gaps between students belonging to the same groups (e.g. boy or girl peer groups) by lifting low performing students more than high performing students. This is a well-known effect of social interac-

[8] Partially connected students refer to those who share some but not all demographic characteristics, and indirectly connected students refer to those who don't share any demographic characteristics but have common classmates who share some characteristics with them.

tions, and is a common feature of group interaction models. A second and somewhat unique feature of our *partitioned* group interaction model is that peer effects may increase gaps between students in different demographic groups by limiting direct classroom spillovers across group boundaries. That is, when boys interact more with boys, and girls with girls, spillovers are stronger within demographic groups than between them - potentially exacerbating any existing gender performance gap. This is an important implication of our model: classroom exposure to other demographic groups is important, but facilitating cross-group interaction may be equally important to close cross-group performance gaps.

To further motivate our model, the next section includes a brief literature review of social network studies in education. Section 3 details the data used in this paper and how we construct our sample. Section 4 introduces the model and discusses identification and estimation. Section 5 presents the main empirical results, and Section 6 explores how these effects change as the classroom composition varies. In particular, we explore how the ethnicity network's impact changes as classroom diversity varies. Section 7 concludes. Appendices include technical details about the identification of our model, simulations results examining the finite sample performance of our proposed model, and some robustness checks using alternative specifications.

2 Social Networks in Education

Despite their potential importance, econometric estimation of peer effects in education remains difficult for a number of reasons. The primary challenge in any peer effect analysis is that the 'true' network is almost never observed, and can only be approximated.[9] Empirical specification of networks involves determining the shape of the network (who is connected to whom?) and the strength of the network (the magnitude of the individual connections within the network). Neither of these empirical choices is trivial, which may explain the

[9] An exception may be production networks where worker interactions may be observed. See Horrace et al. (2016) and Horrace et al. (2020) for examples.

varied results in the literature.[10]

Previous work uses a variety of sources to construct peer networks: college roommate (Sacerdote, 2001), classroom seating assignments (Hong and Lee, 2017), squadrons in the US Air Force Academy (Carrell et al., 2009; Carrell et al., 2013), number of shared college classrooms (De Giorgi et al., 2010), student lunch lines (Presler, 2020) among others. However, these approaches use data unique to specific situations that are not readily generalized.

A reasonable alternative to the aforementioned approaches is using shared student characteristics (homophily) as a proxy for the network. Race and gender are the most common shared characteristics we see used (Arcidiacono and Nicholson, 2005, Renna et al., 2008, Lavy et al., 2012, Hsieh and Lin, 2017, Ananat et al., 2018, Billings et al., 2019), but other characteristics such as immigration status (Damm, 2014) also have been considered. Our paper follows this line of research but considers a broader set of homophilous factors simultaneously in explaining peer effects among elementary school students. Specifically, we examine the relative importance of within-classroom networks based on shared neighborhood of residence, bus route, bus stop, native language, country of birth, gender, and ethnicity.

An added benefit of using demographic data to construct peer networks is that it doesn't require network data that are costly to collect such as friendship surveys (e.g., Add Health; Patacchini et al., 2017) or social media data (e.g., Facebook; Mayer and Puller, 2008).

3 Data and Sample

This paper uses longitudinal student-level administrative data on student academic performance, socio-demographic characteristics, school and classroom codes, and program participation, which we link to student bus route and bus stop assignment data.

Student-level demographic data comes from the New York City Department of Education (NYCDOE). These data include socio-demographic characteristics such as gender, race, age,

[10] The survey of the empirical effects in Sacerdote (2001) highlights this variety. More recently, the survey Paloyo (2020) confirms that this is still true, although we have progressed in our understanding of these effects - particularly in the difficulty entailed in implementing policy based on peer group manipulation.

grade, residential neighborhood, country of birth, primary language spoken at home, an indicator of eligibility for free or reduced-price lunch, and an indicator for students with disabilities. Data also include class assignment as well as both current year and lagged mathematics and reading test scores. We compile this data for academic years 2013-2015.

Transportation assignment data are provided by the NYCDOE Office of Pupil Transportation (OPT). Data include whether a student is assigned a bus, bus route number, bus stop location (latitude and longitude), and bus pickup time all at the student level.

Our sample consists of students in grades 4-5 in general education classrooms where at least two students are assigned a bus. Table 1 shows summary statistics for our sample. Notice that the sample is whiter, less poor, and skewed towards Staten Island than the overall population of students in NYC public schools. This is primarily due to our interest in bus riders and the higher bus ridership in Staten Island. What follows is a discussion about our sample and its characteristics.

We exclude students in K-3 because students start taking standardized tests in grade three, and we include a lagged test score in our model. We focus on elementary students for three reasons. First, in middle school and beyond, students tend to switch classes as they move between subjects, but elementary school students typically remain with the same group of students. Second, elementary school students generally do not choose their classes, and so while the assignment method by which students are placed in a class together is unknown (decided by the principal) class assignment is not a result of students choosing their peers. Finally, bus eligibility in NYC is uncommon for students beyond grade six (only in limited schools in Staten Island).[11] Our sample is restricted to classrooms with at least two bus riders (they do not have to share a route). Table 1 shows that bus assignment is approaching 25%, which is much higher than the nearly 10% we see over all NYC schools. There may be

[11]Students between grades three and six are eligible for subsidized (or free) transportation if they live more than a mile from school. Whether the bus is offered as a transportation choice is determined by the school principal. Weinstein et al. (2021) predicts whether a school offers the bus in NYC. They find that the largest predictors of schools offering the bus are when schools are larger, whiter, and in Staten Island or Queens. This is consistent with how our sample compares to the NYC population of fourth and fifth graders. Schools that do not offer the bus instead offer metro cards.

Table 1: Sample Summary Statistics

Variable		Mean	Std. Dev.	Min	Max
	Class Size	27.8	3.263	20.0	38.0
Test Scores:					
	Math Z-Score	0.414	0.941	-3.748	2.932
	Reading Z-Score	0.406	0.927	-4.757	3.578
	Math Lag Z-Score	0.422	0.924	-5.189	3.838
	Reading Lag Z-Score	0.425	0.894	-6.911	5.573
Student Characteristics:					
	Female	0.505	0.500	0	1
	Ever FRPL	0.731	0.444	0	1
	Age (months)	120.9	7.746	99.5	163.6
	Assigned Bus	0.245	0.430	0	1
	Fourth Grade	0.520	0.500	0	1
	Fifth Grade	0.480	0.500	0	1
Ethnicity:					
	Asian/Other	0.237	0.425	0	1
	Hispanic	0.250	0.433	0	1
	Black	0.164	0.370	0	1
	White	0.349	0.477	0	1
Borough:					
	Manhattan	0.075	0.264	0	1
	Bronx	0.115	0.319	0	1
	Brooklyn	0.160	0.367	0	1
	Queens	0.356	0.479	0	1
	Staten Island	0.294	0.455	0	1
Counts:					
	N Students	55,767		Bus Stops	4,666
	Classrooms	2,181		Bus Routes	1,733
	Census Tracts	1,967		Zip Codes	182

Mean class size is calculated from the pool of classrooms (2,181) such that a classroom in 2014 is distinct from a classroom in 2015 (even if the physical location is identical). All other means are calculated using the pool of 55,767 students.

structural differences in schools that serve more than our target grades (such as K-12 or K-8 schools), so we omit schools serving students above grade five.[12] We limit our analysis to the general education population.[13] Because we are concerned that some classrooms which appear small in our sample may be Integrated Co-Teaching (ICT) classrooms, we remove students in classrooms of less than 20 students (5.8%).

Table 2: Language and Country of Birth Frequency in Sample

Top Languages Spoken at Home		Top Countries of Birth	
	Percent		Percent
Albanian	0.72	Bangladesh	0.68
Arabic	1.19	China	1.13
Bengali	1.58	Dominican Republic	0.52
Chinese (any)	8.44	Ecuador	0.13
English	64.98	Guyana	0.17
Haitian Creole	0.53	Jamaica	0.34
Korean	1.89	Mexico	0.27
Russian	3.44	Pakistan	0.35
Spanish	11.56	Russia	0.22
Urdu	0.97	Trinidad & Tobago	0.04
Other	4.71	USA	91.82
		Other	4.33
Total	100	Total	100

Percentages are caclulated based on the sample of 55,767 student-year observations.

Demographic information includes the language spoken at home and the country of birth. Most students report English as the primary language spoken at home (58.9%) and were born in the US (89%). We construct a list of the most common languages spoken at home and most common countries of birth and utilize this subset of languages in constructing our networks. Table 2 shows which languages and countries are most common and the percentage of the

[12]For example, there may be older students who ride the bus with the students in our sample. While our approach ignores these out-of-class peer effects, having students on the bus whom we do not observe dilutes the chance for students who ride the same bus to be in the same class. This can be systematically different for portions of the sample, and if the group of students choosing to attend K-8 schools is different from those attending K-5 schools, we introduce bias into our peer effect estimates by making the weighing matrix a function of school type (and the differences in student characteristics that comes with that). We avoid this issue by ignoring schools serving higher than grade five.

[13]Specifically, we remove full-time special education students, students who attend a fully special education school, and students who ride a special education bus.

sample each group represents.[14]

4 Empirical Model

4.1 Peer Networks Based on Homophily

This paper constructs peer networks based on the idea that students are drawn to other students with similar observable demographic characteristics (homophily). A number of papers have studied homophily, but typically only a few demographic characteristics are considered individually. Most commonly considered are gender and ethnicity.[15] Our empirical model expands on gender and ethnicity and simultaneously considers a broad set of factors in explaining peer effects among New York City elementary school students.

In our model, student i is the unit of observation and the classroom is the environment in which students interact and perform academically. Therefore, social interactions with students in other classrooms are considered negligible. As such, our model is well-suited for estimating peer effects in elementary schools, where students spend most instructional time in their homeroom.[16] Let the number of students in each of $c = 1, ..., C$ classrooms be n_c, so $\sum_{c=1}^{C} n_c = n$ is the total number of students in the sample or universe. Suppose there are K ways to demographically partition students in a classroom (i.e., gender, ethnicity, etc.).

For ease of exposition, we first consider a simple case of two demographic partitions, gender and ethnicity, (i.e., $K = 2$) and then present a generalized model with K partitions.

[14] As we shall see in Section 4, when constructing peer networks based on homophily, we exclude 'other' categories. That is, students in these categories are not connected to one another and are isolated from the network. For example, a student who speaks French at home does not share a home language connection with a student who speaks Indonesian.

[15] Most studies find that peer effects are stronger within gender or racial groups than across them. See Soetevent and Kooreman (2007), Fruehwirth (2013), Nakajima (2007), Mayer and Puller (2008), Hsieh and Lin (2017), Xu and Fan (2018), among others.

[16] Thus, an implicit assumption in our model is that peer effects occur only in the classroom. This assumption is not uncommon in the literature (e.g. Fruehwirth, 2013), and in our context of elementary education, it may be reasonable to assume the interactions beyond the classroom don't create academic spillovers in appreciable ways. Burke and Sass (2013) show that classroom peers are much more influential than cohort-level peers. This is intuitive, as we should expect classmates who have opportunities to interact during the entire school day to be much more influential on one another than schoolmates with minimal opportunities for interaction.

We assume there are only two genders, boy and girl, and four ethnicity groups, Black, White, Hispanic and Asian. To construct gender and ethnicity networks, we endow each student i in classroom c with scalar values for two unordered categorical variables that determine his/her gender and ethinicity, $h_{g,ic} \in \{1,2\}$ and $h_{e,ic} \in \{1,2,3,4\}$, where g and e stand for gender and ethnicity, respectively; $h_{g,ic} = 1$ indicates student i is a boy; $h_{g,ic} = 2$ indicates a girl; and $h_{e,ic} = 1, 2, 3, 4$ indicate a Black, a White, a Hispanic and an Asian student, respectively. We hereafter refer to these variables as partition variables. In this example, each student belongs to one of the two gender groups and one of the four ethnicity groups.

Then, our gender network matrix (before row-normalization) for classroom c, $\mathbf{W}^*_{g,c}$, is specified as $\left[\mathbf{W}^*_{g,c}\right]_{ij} = w^*_{g,ijc} = \mathbf{1}(h_{g,ic} = h_{g,jc})$ for $i \neq j$ and zero otherwise,[17] where $\mathbf{1}(\cdot)$ is an indicator function that equals one if the argument is true and zero otherwise. The ethnicity network can be constructed similarly. For example, consider a classroom c consisting of five students $\{s_1, s_2, s_3, s_4, s_5\}$, where student s_1 is a Hispanic girl; s_2 is a Hispanic boy; s_3 is an Asian boy; s_4 is an Asian girl; and s_5 is an Asian girl. Then, $\mathbf{W}^*_{g,c}$ and $\mathbf{W}^*_{e,c}$ are given by

$$\mathbf{W}^*_{g,c} = \begin{array}{c} \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \end{array} \qquad \mathbf{W}^*_{e,c} = \begin{array}{c} \begin{array}{ccccc} s_1 & s_2 & s_3 & s_4 & s_5 \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \end{array} \begin{array}{c} \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5. \end{array}$$

In this example, student s_1 is not connected to student s_2 through the gender network, but is connected through the ethnicity network. Also, student s_1 and s_3 don't share any demographic characteristics, so they are not directly connected through the networks. However, as we shall see in the next section, they are *indirectly* connected through other students who share gender or ethnicity with them (e.g., student s_2 shares ethnicity with s_1 and gender

[17]Note that we are using 'exclusive averaging' where individual students are not their own peers. So, the diagonal entries of our network matrices are all zero (i.e. zero-diagonal matrices).

with s_3), and their outcomes can be affected by each other through this intermediating group of students. These partial and indirect connections exist as long as the gender and ethnicity partitions divide students differently. Next section discusses how this variation in partition structures identifies our model.

Let $\mathbf{W}_{g,c}$ be the row-normalized version of $\mathbf{W}^*_{g,c}$. That is, each row of $\mathbf{W}_{g,c}$ sums to 1,[18] then $\mathbf{W}_{g,c}\mathbf{y}_c$ is a vector of mean outcomes of the students with whom each student share gender, where $\mathbf{y}_c$ is an $n_c \times 1$ outcome vector. $\mathbf{W}_{e,c}$ is similarly defined. In this manner, we can construct K homophily networks. We hereafter index each of the demographic partitions/networks by $k = 1, ..., K$, and denote the number of possible categories within k^{th} partition by R_k (e.g., $R_k = 2$ for gender partition).[19] Accordingly, for demographic partition k, $h_{k,ic} = \{1, ..., R_k\}$ and $\mathbf{W}_{k,c}$ indicate the partition variable and homophily network matrix, respectively.

Given these K homophilous weighting matrices, one may sum them to a single, aggregate network such that $\mathbf{W}^*_c = \mathbf{W}^*_{1,c} + ... + \mathbf{W}^*_{K,c}$, row-normalize it, and estimate a single peer effect on the aggregate network. This is the approach of De Giorgi et al. (2010) where k indexes different sections of nine courses to which college students are randomly assigned. This approach implicitly assumes equal weighting for all partitions (i.e., each course is equally important in creating social connections and enhancing student productivity), and it appears to be an assumption of convenience.

However, our goal is to determine the relative importance of these partition networks in explaining peer interactions. Therefore, we do not aggregate the networks and consider a more general group interaction model such that

$$\mathbf{y}_c = \sum_{k=1}^{K} \lambda_k \mathbf{W}_{k,c}\mathbf{y}_c + \mathbf{X}_c\beta + \sum_{k=1}^{K} \mathbf{W}_{k,c}\mathbf{Z}_c\theta_{\mathbf{k}} + \delta_c \iota_{n_c} + \mathbf{u}_c. \tag{1}$$

[18]The row-normalized gender weights matrix $\mathbf{W}_{g,c}$ is given by $\mathbf{W}_{g,c} = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}$.

[19]Since each student belongs to one of the groups in each partition, each student belongs to K groups, and there exist $\sum_{k=1}^{K} R_k$ total groups in the model.

$\mathbf{X}_c$ is an $n_c \times p$ matrix of exogenous variables consisting of a matrix of variables $\mathbf{Z}_c$ that vary within demographic groups and a matrix of dummy variables $\mathbf{D}_c$ generated from the partition variables $h_{k,ic}$ for all k. That is, $\mathbf{D}_c$ consists of $\sum_{k=1}^{K}(R_k - 1)$ columns associated with $\sum_{k=1}^{K} R_k$ demographic groups at our disposal. Thus, $\mathbf{X}_c = [\mathbf{Z}_c, \mathbf{D}_c]$ and $\beta = [\beta_1', \beta_2']'$ where β_1 and β_2 are coefficient vectors on $\mathbf{Z}_c$ and $\mathbf{D}_c$, respectively. ι_{n_c} is an $n_c \times 1$ vector of ones, and $\mathbf{u}_c$ is an error vector satisfying $E(\mathbf{u}_c|\mathbf{X}_c, \delta_c, \{\mathbf{W}_{k,c}\}_{k=1,\dots,K}) = 0$.[20]

Parameters λ_k and θ_k capture endogenous and exogenous (contextual) peer effects, respectively, associated with k^{th} homophily network, and δ_c captures correlated effects among students in classroom c (Manski, 1993).[21] Endogenous peer effects measure the extent to which one's outcome is influenced by her peers' outcomes; exogenous peer effects measure the extent to which one's outcome is influenced by her peers' characteristics; and correlated effects, in our context, may arise from non-random classroom assignment or common environmental factors faced by students in the same classroom (e.g. teacher quality). All three effects cause correlations in the outcome between students in a classroom, but only the first two are the result of peer/social interactions.

Endogenous peer effects are of particular interest since these effects produce a social multiplier. This can be seen from equation (1), where any changes in student i's outcome affect her peers' outcomes through the endogenous peer effect term, but the change in her peers' outcomes affects her peers' peers' outcomes and also i's outcome again. These feedback and chain effects imply that the effect of any policy targeting a subpopulation can be socially shared and multiplied if such an indirect social effect exists. However, identifying the three effects separately in the model is challenging since the endogenous peer effect is some function of the other effects, which can be seen from the reduced form of equation (1) (see equation

[20] A scalar form of equation (1) is given by

$$y_{ic} = \sum_{k=1}^{K} \sum_{j=1,j\neq i}^{n_c} \lambda_k w_{k,ijc} y_{jc} + \mathbf{x}_{ic}' \beta + \sum_{k=1}^{K} \sum_{j=1,j\neq i}^{n_c} w_{k,ijc} \mathbf{z}_{jc}' \theta_k + \delta_c + u_{ic}.$$

[21] Note that there is perfect multicolinearity among the RHS variables in equation (1) if we include the partition dummies $\mathbf{D}_c$ in $\mathbf{Z}_c$ of the exogenous peer effect term.

(A.2) in Appendix A). Therefore, there may be strong correlations between the terms, which raises an identification issue, referred to as *reflection problem* in Manski (1993). Various identification conditions/restrictions have been studied in the literature to separately identify the effects. In the next section we build on existing identification results to understand how our group interaction model is identified.

4.2 Sources of Identification and Potential Issues

Identification of our model hinges on variations in partition structures and group sizes.[22] In our model, students experience peer effects from different demographic groups depending on their demographic characteristics. Therefore, the distribution of student outcomes in a classroom is affected by how each demographic partition divides students into groups and how groups of one partition intersect with the others. Importantly, we show below that variation in partition structures generates exclusion restrictions similar to those by "transitive triads" (Bramoullé et al., 2009) and "partially overlapping groups" (Laschever, 2013; De Giorgi et al., 2010) in the literature. Therefore, compared to the group interaction model of Lee (2007), which solely relies on group size variation for identification, our identification is less mechanical.

Consider again the case with two partitions, each with two groups in the partition, gender (male and female) and ethnicity (Hispanic and Asian). Provided that the exogenous variables $\mathbf{z}_{ic}$ in $\mathbf{Z}_c$ of equation (1) vary sufficiently across students, it is easy to see that the exogenous gender peer effect term (i.e., $\mathbf{W}_{g,c}\mathbf{Z}_c$) and the exogenous ethnicity peer effect term (i.e., $\mathbf{W}_{e,c}\mathbf{Z}_c$) contain different variations for each student as long as the gender and ethnicity partitions divide students differently (i.e., $\mathbf{W}_{g,c} \neq \mathbf{W}_{e,c}$), so they can be separately identified. For instance, in the case of Hispanic boys, their exogenous gender effect terms contain the characteristics of Asian boys but not that of Hispanic or Asian girls, whereas their exogenous

[22]The structure of each demographic partition determines the structure of the associated homophily network. So, the terms "variation in partition structure" and "variation in network structures" are interchangeable. Also, partition structure determines the relative sizes of the groups within each partition, but the absolute sizes of the groups are ultimately determined by classroom sizes.

ethnicity effect terms contain that of Hispanic girls but not that of Asian boys or girls.

Also, observe that in the model without endogenous social effects (i.e., $\lambda = 0$), there is no way for Hispanic boys to affect the outcomes of Asian girls since they don't share any demographic characteristics. This is true for any pair of groups that don't share any demographic characteristics. However, if endogenous *gender* peer effects exist, Hispanic boys can affect Asian girls through *Hispanic girls* such that characteristics of Hispanic boys affect the outcomes of Hispanic girls, which in turn affect the outcomes of Asian girls. Similarly, if endogenous *ethnicity* peer effects exist, *Asian boys* can intermediate such indirect interaction. This implies that, as long as there exists such indirectly connected groups, which depends on the structures of the partitions, endogenous social effects generate variations different from those by exogenous effects and thus the two effects can be separately identified.

These are exclusion restrictions generated by the partition structures in our model, similar to the restrictions by "transitive triads" (Bramoullé et al., 2009) and "partially overlapping groups" (Laschever, 2013; De Giorgi et al., 2010). De Giorgi et al. (2010) use the idea of partially overlapping groups to determine network weights based on how frequently two students met in a series of classes. Their model estimates a single peer effect parameter using an *aggregated* network, which uses variation in network weights at the individual level as exclusion restrictions. Laschever (2013) considers a similar multiple reference group framework, but focuses on estimating peer effects through one network, while using the other networks as exclusion restrictions.

In addition to the variation in partition structures, variation in group sizes also play a role in the identification of our model. Indeed, our model can be seen as a generalization of the group interaction model of Lee (2007), who considers a linear-in-means model with group fixed effects where individuals interact in groups with equal intensities. Lee's model is a special case of our model where $K = 1$ and and $R_k = 1$. That is, each class is completely homogeneous (a single partition with a single demographic group). He shows that the model is identified if there is sufficient variation in classroom sizes. Intuitively, under exclusive averaging (i.e., individuals are not their own peers), better students have worse peers, which reduces the dispersion in outcomes and this dispersion reduction decreases with group size in the model (Bramoullé et al., 2020). Therefore, if there is sufficient variation in group sizes, peer effects can be identified by exploiting such an impact of group size on outcome dispersion.

The same identification mechanism works for our model, but our model partitions classmates further into multiple demographic groups, which creates much richer variation in group sizes that helps identification. Lee (2007) also shows the rate of convergence of the peer effect estimator is slower when average group size is large relative to the number of groups. Our demographic groups within a classroom are generally small, so our estimates may be more precise than a single classroom-level estimate. Identification based on group size variation may be seen as mechanical, but it can help identification when the variation of partition structure is weak.

Appendix A provides more technical details about the identification of our model using the identification results from Boucher et al. (2014) and Bramoullé et al. (2009). Also, Appendix B includes simulations to examine whether there are sufficient variations in the group sizes and partition structures across classrooms in the NYC data, and the results suggest that the current empirical distributions of the group sizes and partition structures are sufficient for consistent estimation of peer effects.

A potential concern in our identification strategy is that students' outcomes may be correlated even without peer effects simply because similar students were assigned into the same classroom and also they were taught in the same environment (e.g. teacher). These types of confounders are referred to as correlated effects in the literature.[23] Particularly, correlations due to non-random assignment can be seen as network endogeneity in our context since networks are formed based on classroom assignment. The usual econometric treatments for such correlated effects are using fixed effects, but we note that in the teacher quality

[23] Bramoullé et al. (2020) survey the recent literature on correlated effects and network endogeneity in peer effect models.

literature, controlling for lagged student performance appears sufficient to mitigate this type of selection bias (Kane et al., 2013; Chetty et al., 2014). Therefore, we control for past performance (as well as both classroom and network fixed effects) to mitigate this sorting issue Repetitive: delete this phrase?. We also examine the randomness of classroom assignment for our sample using a series of multinomial logit analyses, which we report and discuss in Appendix C.1 and show no systematic patterns in classroom assignment associated with demographic variables. This is not direct evidence for random assignment, but indicates that classroom assignment is not a function of the associated demographic partitions.

Another potential issue is that there may be unobserved networks that are correlated with our demographic networks. In particular, if these missing links are associated with our demographic networks differentially, our rankings of peer effects across demographic partitions will be biased. One potentially important network might be student performance. Some papers in the peer effect literature suggest that students may interact along performance levels (e.g., Carrell et al., 2013). Therefore, we include a peer network based on past performance as a control and check the sensitivity of our estimates.[24] The results reported in Appendix C.2 imply that our estimates are largely robust to this network.[25]

4.3 Empirical model

There are a few differences between equation (1) and the model that we ultimately estimate. First, we use a panel of data, so we employ both cross-sectional and time variations in peer networks and covariates (i.e. $\mathbf{W}_{k,ct}$ and $\mathbf{X}_{ct}$) to estimate model parameters. This also allows us to add a one period lag of the dependent variable, $\mathbf{y}_{c,t-1}$, on the right-hand side of the

[24] To construct the past performance network, we rank students in each classroom based on previous year's test scores and then divide them into groups based on quantile.

[25] Appendix B includes a simulation experiment that generates data using a standard peer effect model with a single friendship network, where friendship is formed based on some shared demographic characteristics, and then applies our homophily network model to the data. We find that the rankings of demographic peer effect estimates from our model reflect the relative importance of each demographic characteristic in the friendship formation. This suggests a decomposition or approximation feature of our model: when the true links are unobserved, our estimates decompose the peer interactions into the portions associated with each demographic characteristic/partition and capture their importance in explaining the interactions.

model to control for student's persistent heterogeneity in test scores. Due to the addition of the time dimension, classroom fixed-effects, δ_c, are expanded to classroom by time fixed-effects, δ_{ct}.

Second, we have assumed so far that each partition is complete. That is, we have assumed that for each partition, every student must belong to exactly one demographic group in each partition. However, there are groups in some partitions that do not facilitate a connection between students. For example, some students ride the bus while others do not. Those that ride the bus together may form stronger in-classroom connections. These connections may be the result of friendships strengthened while traveling on the bus together, or they may reflect homophily based on shared experiences due to shared geography. Neither of these reasons for increased interaction occur for students who do not ride the bus. In this case, the weighting matrix for bus ridership may include a block of zeros for students who do not have connections with other students in this partition (i.e. those who do not ride the bus are excluded from the bus ridership network). These serve as additional exclusion restrictions that helps identify the model.

Therefore, our empirical model is:

$$\mathbf{y}_{ct} = \sum_{k=1}^{K} \lambda_k \mathbf{W}_{k,ct}\mathbf{y}_{ct} + \mathbf{X}_{ct}\beta + \sum_{k=1}^{K} \mathbf{W}_{k,ct}\mathbf{Z}_{ct}\theta_k + \mathbf{y}_{c,t-1}\gamma + \delta_{ct} \cdot \iota_{n_{ct}} + \mathbf{u}_{ct} \tag{2}$$

where n_{ct} is the number of students in classroom c at time t. We include seven homophily peer networks: shared Gender, shared Ethnicity, shared Language Spoken at Home, shared Country of Birth, shared Bus Route, shared Bus Stop, and shared Residential Neighborhood (census tract or zip code). The covariate vector, $\mathbf{X}_{ct}$, includes: a gender dummy ($Female$), a dummy for 'free or reduced price lunch' ($FRPL$), Age in months, Age^2, a dummy for a bus rider (Bus), three dummies for race ($Asian$, $Black$ and $White$ with $Hispanic$ as the baseline group), and dummies for other indicator variables that are used to construct the weights matrices (e.g. dummies for language spoken at home, country of birth, etc.). The covariate vector for contextual effects, $\mathbf{Z}_{ct}$, includes $FRPL$, Age, Age^2, and lagged outcome

$\mathbf{y}_{c,t-1}$.[26]

We estimate the model using a quasi-maximum likelihood method in a similar fashion as Lee et al. (2010). Estimation details are in Appendix D. The estimation approach requires that the determinant of $\mathbf{I}_{ct} - \sum_{k=1}^{K} \lambda_k \mathbf{W}_{k,ct}$ is strictly positive for any c and t in equation (2) to ensure the log-likelihood function is well defined. When weights matrices are row-normalized (e.g. our case), the condition is satisfied if $\sum_{k=1}^{K} |\lambda_k| < 1$. Therefore, we only consider parameter values satisfying $\sum_{k=1}^{K} |\lambda_k| < 1$ in our estimation procedure. This ultimately implies that not all peer effects of different partitions can be large, but only some of them can. As we are mostly interested in the relative importance of these factors, this does not pose a problem.

When the peer effect estimates are ranked to examine the relative importance of the partitions, the ranking statistics are not deterministic, but contain errors due to sampling variability. We apply the method of 'ranking and selection' to properly infer the most important partitions from the ranked estimates. The procedures select a subset of $\lambda_1, ..., \lambda_K$ that are statistically larger than the others at a pre-specified error rate, $\alpha \in (0, 0.5)$, while accounting for the inherent multiplicity and uncertainty in the ranked estimates, $\hat{\lambda}_k$'s.[27] Let $\lambda_{[K]} \geq \lambda_{[K-1]} \geq ... \geq \lambda_{[1]}$ be the population rankings of peer effects, so $\lambda_{[K]} = \max_{k=1}^{K} \lambda_k$. The rankings of peer effects estimates are similarly denoted as $\hat{\lambda}_{(K)} \geq \hat{\lambda}_{(K-1)} \geq ... \geq \hat{\lambda}_{(1)}$, so $\hat{\lambda}_{(K)} = \max_{k=1}^{K} \hat{\lambda}_k$. Ranking and selection procedures recognize the uncertainty that the partition that is estimated to have j^{th} largest peer effects, (j), may not correspond to the partition that has the j^{th} largest peer effects in the population, $[j]$, in general.

To account for such uncertainty, with assuming (asymptotic) normality of the estimates $\hat{\lambda}_1, ..., \hat{\lambda}_K$ and general variance-covariance structure, the procedures identify a set of peer effect indices, $\zeta \subset \{1, 2, ..., K\}$ that satisfies $Pr\{[K] \in \zeta\} \geq 1 - \alpha$ where $\alpha = 0.05$, typically.

[26] As discussed before, partition dummies are excluded from $\mathbf{Z}_{ct}$ to avoid perfect multicollinearity. Since there are seven homophily weights matrices and four covariates in $\mathbf{Z}_{ct}$, twenty eight contextual effects are estimated in our model. We do not report the contextual effect estimates in our results below to save space (these estimates are available upon request from the authors).

[27] See Horrace and Parmeter (2017), who recently apply ranking and selection to economics journal citation counts to determine a subset of the 'best' journals.

In other words, the procedures estimate the (*minimal*) set of the population indices that includes the *unknown* population index that is associated with the largest parameter value, $[K]$, with probability at least $1 - \alpha$. If the inference is very sharp, ζ may be a singleton, but ζ may include all the indices if the inference is very weak.[28]

5 Main Results

We begin with our results for mathematics and reading scores for fourth and fifth graders in Table 3. The test scores are normalized Z-scores. To save space we do not report the contextual effect results, only the peer effects (λ_k, $k = 1, ..., 7$) and the marginal effects (β) of the covariates ($\mathbf{X}_{ct}$) in equation (2). Table 3 shows the results for four models. Models (1) and (2) use mathematics test score as the outcome, and models (3) and (4) use reading test score as the outcome. The first model for each outcome defines a student's neighborhood as their zip code of residence (models (1) and (3)) and the second model for each outcome defines a student's neighborhood as their census tract of residence (models (2) and (4)). Notice that this change has very little effect on estimates for other variables in the model.

Four of the mathematics score peer effects are significant in column (1), with Ethnicity (0.221), Language (0.096), Gender (0.054), and Neighborhood (0.017) in rank order. Bus Stop, Bus Route, and Country are not statistically significant.[29] In column (2) where we define Neighborhood using census tract rather than zip code, the estimates are very similar, with a slight increase in the point estimate for Neighborhood. The four networks that were significant in model (1) remain significant in model (2), and the insignificant estimates remain

[28] The cardinality of ζ is increasing in K since the inference needs to make more pairwise comparisons as K increases, which makes it harder to distinguish $[K]$ at a fixed error rate, α. This is the concept of 'multiplicity.' To conduct the inference we need critical values drawn from a k-dimensional multivariate normal distribution with covariance structure determined by the hessian of the converged log-likelihood. These critical values were all around 2.5, larger than the usual critical value of 1.96 from a univariate normal distribution and, hence, accounting for the multiplicity.

[29] We should note that we have not included fixed effects for specific bus stops and bus routes due to computational constraints, although we have included an indicator for whether students ride the bus. As a result, these estimates may be biased by correlated effects related to the bus. However, these results are already much weaker than other networks, and the additional controls are unlikely to significantly increase the influence of these networks.

Table 3: Main Demographic Peer Effect Results

	Mathematics				Reading			
	(1)		(2)		(3)		(4)	
Peer Effects (λ_k)								
Ethnicity	0.221	(0.008)	0.221	(0.008)	0.117	(0.009)	0.116	(0.009)
Language	0.096	(0.009)	0.096	(0.009)	0.052	(0.009)	0.052	(0.009)
Gender	0.054	(0.013)	0.054	(0.013)	0.150	(0.012)	0.150	(0.012)
Nbrhd (Tract)			0.018	(0.006)			0.012	(0.006)
Nbrhd (Zip)	0.017	(0.009)			0.005	(0.009)		
Bus Stop	0.020	(0.013)	0.018	(0.013)	0.008	(0.013)	0.006	(0.013)
Bus Route	0.008	(0.012)	0.009	(0.012)	0.028	(0.012)	0.028	(0.012)
Country	0.008	(0.016)	0.008	(0.015)	0.024	(0.017)	0.024	(0.016)
Marginal Effects (β)								
Lag Test Score	0.711	(0.003)	0.709	(0.003)	0.687	(0.003)	0.683	(0.003)
Age	0.016	(0.007)	0.015	(0.007)	0.035	(0.008)	0.034	(0.008)
Age Squared	-0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)
Male	0.007	(0.004)	0.009	(0.004)	-0.047	(0.005)	-0.047	(0.005)
FRPL	-0.076	(0.006)	-0.074	(0.006)	-0.070	(0.007)	-0.068	(0.007)
Bus	-0.021	(0.010)	-0.024	(0.010)	-0.022	(0.010)	-0.024	(0.011)
Asian	0.117	(0.010)	0.112	(0.010)	0.091	(0.011)	0.088	(0.011)
Black	0.011	(0.009)	0.011	(0.009)	-0.011	(0.010)	-0.007	(0.010)
White	0.041	(0.009)	0.036	(0.009)	0.030	(0.009)	0.022	(0.010)
Additional Controls	Yes		Yes		Yes		Yes	
Cont. Soc. Effect	Yes		Yes		Yes		Yes	
Classroom FE	Yes		Yes		Yes		Yes	
Zip Code FE	Yes		No		Yes		No	
Census Tract FE	No		Yes		No		Yes	
Country of Birth FE	Yes		Yes		Yes		Yes	
Lang. at Home FE	Yes		Yes		Yes		Yes	
Observations (n)	55,598		55,598		55,221		55,221	

Models (1) and (3) define the neighborhood using student's residential zip code. Models (2) and (4) define the neighborhood using census tract. Models include 28 contextual social effects, constructed using the lag test score, age, age squared, and FRPL status along with each of the seven networks included in the model, as well as neighborhood and classroom fixed effects. Hispanic is the omitted reference group for ethnicity marginal effects.

insignificant as well.

Turning to the results for Reading Test Z-Scores, even though the sample selection process is identical to that for the mathematics scores, there are some notable differences in peer effects across the two sets of results. First, the mathematics peer effects are typically larger than the reading peer effects except for Gender (compare 0.054 for mathematics to 0.150 for reading) and Bus Route (compare statistically significant 0.028 in reading to insignificant 0.008 in mathematics). Second, the peer effects rankings are about the same for reading and mathematics scores. The biggest difference is that Gender is ranked third in mathematics (0.054), but is ranked first in reading (0.150). Additionally, Bus Route is significant in reading (0.028) and ranks before Neighborhood. Country of birth and Bus Route remain statistically insignificant for both outcomes. Only the census tract definition of Neighborhood is statistically significant in reading. In both outcomes, the estimate is larger for the census tract definition of Neighborhood, suggesting that the census tract is a better representation of an elementary student's neighborhood peer group.[30]

The ranked peer effect estimates contain sampling errors, so we examine if the the rankings of peer effect estimates are statistically significant using the ranking and selection procedures. We find that only Ethnicity is in the subset of the best in case of mathematics test scores while both Ethnicity and Gender are in the best set in the case of reading test scores. This is essentially because the Ethnicity and Gender peer effect estimates are considerably larger than the others, so they remain in the best category even after accounting for sampling errors.

A few remarks follow: first, the three strongest networks are consistent between both outcomes (Ethnicity, Language, and Gender). Second, our estimates imply that social effects are larger in mathematics than in reading.[31] This is consistent with the story that students

[30] While we would like to find the "optimal" neighborhood size, we were unable to push this further with census blocks, due to computational difficulties with the large number of fixed effects that would entail (we incorporate over 2,100 census tracts already). We also had concerns of being underpowered with census blocks (there are over 23,000 unique census blocks associated with our sample). The bus networks may substitute as smaller neighborhood definitions, particularly the Bus Stop network. Our results suggest that this is not a better approximation than census tract, but this could also be attributed to the sparsity of the bus network.

[31] See that more networks exhibit larger coefficients in math than in reading, and also the sum of network estimates is larger in mathematics than in reading.

learn language and reading skills both at home and at school, but mathematics is learned primarily at school. Next, the primary Language spoken at home is strong for both outcomes, but stronger in mathematics than reading. The smaller effect and importance of Language in reading may come from two mechanisms. The first is that schools teach English specifically, so other language groups may play diminished roles. The second is the aforementioned idea that students primarily learn mathematics at school, such that we might expect a smaller coefficient for reading estimates in general. We explore the primary Language spoken at home and Country of birth networks further in Section C.

Additionally, these results suggest that influential peers are generally those that share obvious similarities: Ethnicity, Gender, and Language. Interactions outside the classroom, such as shared Bus Stop, Bus Route, and neighborhood appear less important. A narrower ethnicity group (Country of birth) appears less important. This could be a result of friendship sorting occurring early in the year, making characteristics which are immediately obvious to students more important for friendship formation.

The next panel of Table 3 contains estimates for the marginal effects of our control variables ($\mathbf{X}_{ct}$). For mathematics, being male is associated with increased scores (borderline significant 0.007), while Free or Reduced-Price Lunch (FRPL) status correlates with lower scores (significant -0.076). The reference group Hispanic does worse than the other three groups, although Black is not statistically different. Lag Test Score is the strongest predictor of current-year performance (0.711). Students who ride the bus do worse than those who do not (-0.021), although this may be due to selection. Age (measured in months) is also statistically significant, and we include both age and age-squared in the model. Being older is correlated with better test results, but being too old (possibly due to grade retention) is associated with lower test results.

It is also important to notice the dimensions along which we observe performance gaps. There are significant performance gaps along both ethnicity and gender, which we observe in the marginal effects (β). Because these are also important dimensions of the peer interaction as seen in the peer effects estimates, equation (A.3) implies that these performance gaps are widened by classroom peer effects. Recall that endogenous peer effects mitigate performance gaps between high and low performing students as low performing students are more helped by their higher performing peers. However, they may increase existing gaps between groups if peer interactions exist within the boundaries of groups. Therefore, within ethnicity and gender groups, there may be significant peer effects, and these effects may exacerbate performance gaps between different gender or ethnic groups. Using the point estimates from Table 3, our results suggest that any within-classroom gender test score gap would be increased by 5.71% in mathematics and 17.65% in reading through the classroom peer effects mechanism. Similarly, the black-white test score gap would increase by 28.37% in mathematics and 17.65% in reading. Nonetheless, these negative impacts are mitigated in our model if there is enough variation of students within demographic groups as they "move" across partitions. For example, performance gaps between boys and girl are less affected if there is sufficient variation of ethnicity among boys and girls (and if there are positive spillovers in ethnicity).

Table 4: Peer Effects (λ from Individual Models)

	Mathematics		Reading	
Ethnicity	0.246	(0.008)	0.128	(0.009)
Language	0.156	(0.008)	0.078	(0.009)
Gender	0.056	(0.013)	0.149	(0.012)
Census Tract	0.026	(0.006)	0.016	(0.006)
Zip Code	0.034	(0.009)	0.014	(0.009)
Bus Stop	0.034	(0.011)	0.030	(0.011)
Bus Route	0.026	(0.010)	0.034	(0.010)
Country	0.038	(0.016)	0.040	(0.017)
Observations	55,598		55,221	

Estimates shown are the parameters of interest from the respective single-network models. Models include classroom fixed effects, four contextual effects, lagged test score, age, age squared, and fixed effects for gender, poverty, ethnicity, census tract, country of birth, language spoken at home, and whether student rides the bus. Ethnicity is defined with Hispanic as the reference group.

Table 4 shows peer effects estimates when the model includes only *one* of our seven homophilous peer networks. The goal of this exercise is to show stark differences in peer effect estimates when they are estimated individually.[32] Nearly all estimates are larger and statistically significant, and the rankings of the estimates are quite different than before. This implies the current practice of considering only a few characteristics individually when studying homophily may suffer from omitted variable bias, and it is indeed important to simultaneously control for all networks to tease out the individual contribution of each homophilous factor on peer interaction.

6 Classroom Composition: An Exploration of Diversity

In this section, we ask how the effect of the Ethnicity network changes as the classroom becomes more ethnically diverse. Does the ethnicity network become more or less important in a more diverse setting? Does classroom diversity affect the influence of networks correlated with ethnicity? Do networks orthogonal to ethnicity substitute for the ethnicity network? These are questions we will explore.

For this analysis, we construct a measure of diversity. Notice that there are four groups of interest in our Ethnicity network: Hispanic, Black, White, and Asian/other. To deal with more than two groups, we use Theil's entropy index (Theil, 1972, White, 1986), which we normalize such that values fall between zero and one:

$$h_c = -\frac{\sum_{k=1}^{K} p_{ck} \ln(p_{ck})}{\ln(K)} \tag{3}$$

where p_{ck} is the proportion of students in classroom c who are members of that ethnicity k. h_c is the normalized entropy index for classroom c, and we will call it the diversity index. The index is maximized when there is equal representation of all groups (i.e. $\forall\ k$, $p_{ck} = \frac{1}{K}$), and minimized at zero when only one group is represented (i.e. $\exists\ k$, $p_{ck} = 1$). We calculate this diversity index for each classroom in our sample, and then split the data based on diversity

[32]These models include the same contextual effects and marginal effects as in Table 3.

quantile.

Table 5 shows the results from the subsamples above and below the median diversity index. In both outcomes we see that the Ethnicity network is more influential in the more diverse environment. The difference is substantial. In mathematics, the peer effect drops from 0.233 to 0.192 (percent difference of 19%) and in reading it drops from 0.130 to 0.092 (34%). Table 6: Panel A divides the mathematics sample into quartiles by diversity. Column (1) is the most diverse, and column (4) contains the least diverse classrooms. Notice that columns (1)-(3) have similar estimates for the Ethnicity network, but column (4) is much smaller. Comparing column (1) to column (4) we see a drop from 0.224 to 0.153 (38%). This pattern is consistent in Panel B, for reading. Columns (1)-(3) have similar estimates of the Ethnicity peer effect, but the estimate in column (4) is starkly different. Again comparing columns (1) and (4) we see a drop from 0.128 to 0.058 (75%).

There may be two potential mechanisms at play, which we discuss with an example of two groups for simplicity: a majority and minority. If the two groups are equal sized, there is no majority, and diversity is maximized. This is analogous to column (1) in Table 6. On the other hand, column (4) occurs when the minority group is quite small, so that there are few to no peers from their group. In this case, the majority group is large enough that their group identify is largely irrelevant as a tool for subdividing the classroom and selecting friends. However, minority students may be isolated from the classroom and their group peer effect is near-zero in such a situation. These two channels are likely to lead to the full drop that we observe: majority students find ethnicity to be a less influential network when most of their peers share the same ethnicity,[33] and minority students become isolated. Despite the difference in context, these results may help to explain why Arcidiacono and Nicholson (2005) finds peer effects along gender lines but not ethnicity lines in US medical schools, as medical schools are known to lack diversity (Lett et al., 2019).

[33] This story is consistent with the findings of Mayer and Puller (2008). They find that minority students in particular are more likely to make connections with other minorities at their university. This effect is larger when the minority group makes up a smaller portion of the student body.

Table 5: Partitioned by classroom ethnic diversity

	Mathematics				Reading			
	More Diverse		Less Diverse		More Diverse		Less Diverse	
Peer Effects (λ_k)								
Ethnicity	**0.233**	**(0.010)**	**0.192**	**(0.013)**	**0.130**	**(0.011)**	**0.092**	**(0.015)**
Language	0.105	(0.011)	0.083	(0.014)	0.057	(0.012)	0.043	(0.014)
Gender	0.047	(0.019)	0.060	(0.018)	0.133	(0.018)	0.167	(0.017)
Neighborhood	0.028	(0.012)	0.005	(0.012)	0.002	(0.012)	0.008	(0.013)
Bus Stop	0.052	(0.020)	0.001	(0.016)	0.004	(0.021)	0.010	(0.017)
Bus Route	-0.002	(0.018)	0.013	(0.015)	0.029	(0.018)	0.027	(0.016)
Country	0.000	(0.021)	0.017	(0.022)	0.043	(0.023)	0.002	(0.024)
Observations (n)	27,451		28,147		27,289		27,932	

Models include 28 contextual social effects, constructed using the lag test score, age, age squared, and FRPL status along with each of the seven networks included in the model, as well as neighborhood and classroom fixed effects. Hispanic is the omitted reference group for ethnicity marginal effects.

Table 6: Quartile Partitions of Classroom Ethnic Diversity

Panel A: Mathematics	(1)		(2)		(3)		(4)	
Peer Effects (λ_k)								
Ethnicity	**0.224**	**(0.013)**	**0.242**	**(0.014)**	**0.213**	**(0.014)**	**0.153**	**(0.023)**
Language	0.076	(0.016)	0.129	(0.015)	0.103	(0.015)	0.045	(0.023)
Gender	0.051	(0.026)	0.043	(0.026)	0.101	(0.019)	0.014	(0.028)
Neighborhood	0.022	(0.017)	0.033	(0.016)	0.003	(0.013)	0.007	(0.019)
Bus Stop	0.086	(0.029)	0.028	(0.027)	0.005	(0.019)	-0.002	(0.023)
Bus Route	-0.033	(0.025)	0.018	(0.024)	0.006	(0.018)	0.020	(0.022)
Country	0.038	(0.029)	-0.032	(0.030)	0.011	(0.023)	0.029	(0.035)
Observations (n)	13,520		13,931		14,019		14,128	
Panel B: Reading								
Peer Effects (λ_k)								
Ethnicity	**0.128**	**(0.015)**	**0.133**	**(0.016)**	**0.112**	**(0.018)**	**0.058**	**(0.025)**
Language	0.045	(0.017)	0.069	(0.016)	0.055	(0.018)	0.020	(0.024)
Gender	0.147	(0.024)	0.118	(0.024)	0.175	(0.023)	0.158	(0.024)
Neighborhood	0.017	(0.017)	-0.012	(0.017)	0.013	(0.017)	0.003	(0.019)
Bus Stop	0.018	(0.030)	-0.007	(0.028)	0.016	(0.024)	0.005	(0.023)
Bus Route	0.008	(0.025)	0.046	(0.025)	0.012	(0.023)	0.041	(0.022)
Country	0.059	(0.031)	0.031	(0.031)	0.003	(0.032)	0.002	(0.036)
Observations (n)	13,458		13,831		13,928		14,004	

Models include 28 contextual social effects, constructed using the lag test score, age, age squared, and FRPL status along with each of the seven networks included in the model, as well as neighborhood and classroom fixed effects. Hispanic is the omitted reference group for ethnicity marginal effects.

It is important to note that this exploration is correlational.[34] We have not randomly assigned students into diverse classrooms, and school diversity is likely tied to neighborhood choices and other endogenous factors. Nonetheless, these results are striking and suggest that Ethnicity (and possibly other networks) retain their importance over most of the distribution, but drop off in importance when diversity is low. There are two questions this exercises raises which are beyond the scope of this paper. The first is whether unrelated networks are substitutes for one another. That is, if one network (ex: Ethnicity) is not diverse, does the importance of another group (ex: Gender) increase? [35] We see related behavior in Carrell et al. (2013), where USAFA students from top and bottom ability quartiles were placed in squadrons without the middle performers. This decrease in diversity lead to more within-squadron segregation along ability lines. Second, do different groups within the network place differing values on membership to their group? Hsieh and Lin (2017) finds that middle and high school females are more affected by their peers than their male counterparts. Similarly, they find that white students are more affected by their peers than other racial groups. Presler (2022) looks at the NYC context and finds that white students are less influenced by peer BMI than other ethnicities.

It is useful to look at the story told by Table C.5. In Table C.5, we see that primary Language spoken at home is much more important in Queens - the borough that has the highest proportion of non-English speakers as well as the largest borough in our sample.[36] Coupled with the findings in Table 6, the change in importance that we see in primary Language suggests a similar pattern - diversity matters for network importance.

[34]That said, we see some similar patterns to those presented here when we split by borough, which can be thought of as causal (attending a school outside a student's borough of residence is very rare).

[35]The results in Table 5 are consistent with the idea that Ethnicity and Gender are substitutes based on above and below the median diversity level. However, this relationship is less clear in Table 6 when broken into quartiles based on diversity. The pattern shown in Table C.5 is similarly noisy. Further exploration of this question is desirable.

[36]This is supported further in the Appendix. Table C.3 shows that when we remove connections based on English as a primary language and students born in the US, the primary Language spoken at home and Country of birth networks increase in importance. This suggests the language has higher salience when there is diversity in the network - just as we see in this example in Queens.

7 Conclusion

In this paper, we find that shared ethnicity, gender, and language spoken at home networks are important sources of spillovers for mathematics and reading test scores in NYC elementary schools. This supports the assumption made by much of the literature that gender and ethnicity networks are important. When modeled with these characteristics, neighborhood, bus stop, bus route, and country of birth appear less important for academic peer effects. Even though we find no significance effect, to our knowledge, this is the first study to explore the importance of bus peers in the classroom. These findings may increase our understanding of network formation within the classroom and our ability to predict group behavior after a change in the network.[37]

In the standard social interaction model, endogenous spillovers mitigate performance gaps by lifting the performance of low performing students more than the performance of high performing students. In this paper, we show that if such spillovers work within the boundaries of groups, the performance gap between students in different groups increases (i.e. increased gender gap or racial disparity in academic performance). Since ethnicity is an important within-classroom network for social spillovers, this implies that known performance gaps, such as the racial performance gap, may be exacerbated even in diverse classroom settings.[38]

We also show that altering the make-up of the classroom changes the impact of these networks. In particular, we show that low ethnic diversity is correlated with low impact for that network. This is likely due to the irrelevance of the network for the majority group, but the negative impact of isolation on the minority group may also contribute to the effect.

[37]Consider the experiment in Carrell et al., 2013, where the authors implicitly assumed that students would form friendships in the same manner before and after the policy change. This proved to be a faulty assumption, and successful implementation of policy attempting to harness peer effects in the classroom will likely include an understanding of how within-classroom networks will change due to the policy change. This involves understanding which groups may be influential and how influential groups affect the outcome. This paper contributes to our understanding of both issues.

[38]This does not imply that classroom diversity is not useful. Our estimates point out that the existence of these within-classroom partitions may have a negative effect on performance gaps relative to a model without students sorting along these lines, and highlight the importance of facilitating cross-group interaction as a means of mitigating these gaps.

Further research is needed to disentangle these mechanisms. We also provide suggestive evidence that this is true for diversity in other networks, such as primary language spoken at home.

Lastly, note that the results we present in this paper are multiplicative peer effects which control for contextual effects. This means that interventions targeting a subset of students in the classroom may produce spillovers to those not targeted by the intervention. Our results inform which groups receive the strongest effects of these spillovers in the presence of groups. For example, a program providing additional mathematics tutoring will produce the strongest spillovers to students sharing ethnicity with the tutored students; extra tutoring in reading will result in the strongest spillovers for students sharing either gender or ethnicity. Students who share a primary language spoken at home will also see important spillovers for both outcomes. This type of information should be useful for designing optimal interventions not only in education but also across a variety of fields, and we provide a roadmap for how to utilize demographic data to obtain such useful information.

Appendix A Source of Identification

This appendix provide more technical details about the identification of our model. To understand how our model is identified, we first apply the within transformation to equation (1) in the main text to remove the correlated effects. That is, we premultiply it by $\mathbf{Q}_c = \mathbf{I}_{n_c} - \iota_{n_c}\iota_{n_c}'/n_c$, where $\mathbf{I}_{n_c}$ is the identity matrix of dimension n_c, such that

$$\begin{aligned}\mathbf{Q}_c\mathbf{y}_c &= \mathbf{Q}_c\sum_{k=1}^{K}\lambda_k\mathbf{W}_{k,c}\mathbf{y}_c + \mathbf{Q}_c\mathbf{X}_c\beta + \mathbf{Q}_c\sum_{k=1}^{K}\mathbf{W}_{k,c}\mathbf{Z}_c\theta_k + \mathbf{Q}_c\mathbf{u}_c \\ &= \sum_{k=1}^{K}\lambda_k\mathbf{W}_{k,c}\mathbf{Q}_c\mathbf{y}_c + \mathbf{Q}_c\mathbf{X}_c\beta + \sum_{k=1}^{K}\mathbf{W}_{k,c}\mathbf{Q}_c\mathbf{Z}_c\theta_k + \mathbf{Q}_c\mathbf{u}_c, \qquad \text{(A.1)}\end{aligned}$$

where the second equality is because $\mathbf{W}_c$ is row-normalized and symmetric. Let $\mathbf{y}_c^* = \mathbf{Q}_c\mathbf{y}_c$, and define $\mathbf{X}_c^*, \mathbf{Z}_c^*$, and $\mathbf{u}_c^*$ similarly. Note that the typical element of $\mathbf{y}_c^*$ is $y_{ic} - \bar{y}_c$, where $\bar{y}_c = \frac{1}{n_c}\sum_{i=1}^{n_c} y_{ic}$ (it is same for $\mathbf{X}_c^*, \mathbf{Z}_c^*$, and $\mathbf{u}_c^*$). We assume $\mathbf{I}_{n_c} - \sum_{k=1}^{K}\lambda_k\mathbf{W}_{k,c}$ is invertible[1] so the model is stable and has an equilibrium. Then, the reduced form of equation (A.1) is written as

$$\begin{aligned}\mathbf{y}_c^* &= \left(\mathbf{I}_{n_c} - \sum_{k=1}^{K}\lambda_k\mathbf{W}_{k,c}\right)^{-1}\left[\mathbf{X}_c^*\beta + \sum_{k=1}^{K}\mathbf{W}_{k,c}\mathbf{Z}_c^*\theta_k + \mathbf{u}_c^*\right] \\ &= \left(\mathbf{I}_{n_c} + \sum_{k=1}^{K}\lambda_k\mathbf{W}_{k,c} + \left(\sum_{k=1}^{K}\lambda_k\mathbf{W}_{k,c}\right)^2 + \ldots\right)\left[\mathbf{X}_c^*\beta + \sum_{k=1}^{K}\mathbf{W}_{k,c}\mathbf{Z}_c^*\theta_k + \mathbf{u}_c^*\right] \qquad \text{(A.2)}\end{aligned}$$

In the sequel, let $\mathcal{S}_c(k,r) = \{i : (h_{k,ic} = r)\}$ for $k = 1, ..., K$, and $r = 1, ..., R_k$ (i.e. the index set of students in classroom c whose k^{th} partition variable is r)[2] and $n_c(k,r) = |\mathcal{S}_c(k,r)|$ where $|S|$ is the cardinality of a set S (so, $n_c = \sum_{r=1}^{R_k} n_c(k,r)$). Remember $h_{k,ic}$, defined in the main text, is the k^{th} partition variable which is associated with k^{th} demographic variable that partitions students in a classroom into R_k groups.

[1]Note that a sufficient condition for the invertibility is $\sum_{k=1}^{K}|\lambda_k| < 1$ since $\mathbf{W}_{k,c}$ are row-normalized.

[2]In the example considered in the main text, where a classroom c consists of five students $\{s_1, s_2, s_3, s_4, s_5\}$ and student s_1 is a Hispanic girl; s_2 is a Hispanic boy; s_3 is an Asian boy; s_4 is an Asian girl; and s_5 is an Asian girl, $\mathcal{S}_c(g,1) = \{s_1, s_4, s_5\}, \mathcal{S}_c(g,2) = \{s_2, s_3\}, \mathcal{S}_c(e,1) = \{s_1, s_2\}$, and $\mathcal{S}_c(e,2) = \{s_3, s_4, s_5\}$, where 1 and 2 indicate girl and boy groups, respectively, in gender partition and Hispanic and Asian groups, respectively, in ethnicity partition.

First, we examine how group size variation can be used to identify equation (A.2). For simplicity, we consider the case $K = 1$. When $K = 1$, after some matrix algebra, student i's outcome function (in scalar form) can be written as

$$\begin{aligned} y_{ic} - \bar{y}_c \quad &= \underbrace{\left(\mathbf{z}_{ic} - \bar{\mathbf{z}}_c^{(1,h_{1,ic})}\right)' \left(\frac{\beta_1 - \frac{\theta}{n_c(1,h_{1,ic})-1}}{1 + \frac{\lambda}{n_c(1,h_{1,ic})-1}}\right) + \frac{1}{1 + \frac{\lambda}{n_c(1,h_{1,ic})-1}} \left(u_{ic} - \bar{u}_c^{(1,h_{1,ic})}\right)}_{(*)} \\ &+ \underbrace{\frac{1}{1-\lambda} \left[\left(\bar{\mathbf{d}}_c^{(1,h_{1,ic})} - \bar{\mathbf{d}}_c\right) \beta_2 + \left(\bar{\mathbf{z}}_c^{(1,h_{1,ic})} - \bar{\mathbf{z}}_c\right) (\beta_1 + \theta) + \left(\bar{u}_c^{(1,h_{1,ic})} - \bar{u}_c\right)\right]}_{(**)} \end{aligned} \tag{A.3}$$

where $\bar{\mathbf{z}}_c^{(1,h_{1,ic})}$ is the average of $\mathbf{z}_{ic}$ of the demographic group associated with the first ($K = 1$) demographic partition to which student i belongs. So $\mathbf{z}_{ic} - \bar{\mathbf{z}}_c^{(1,h_{1,ic})}$ represents the difference in attributes between i and her demographic group average, and $\bar{\mathbf{z}}_c^{(1,h_{1,ic})} - \bar{\mathbf{z}}_c$ represents the difference in attribute between her demographic group average and the classroom average. The equation is essentially the same as that derived in Lee (2007), but ours includes the additional terms in (**), which appear because $R_k > 1$ for all k (i.e., because each classroom contains multiple demographic groups within each demographic partition).

Boucher et al. (2014) provide an intuitive explanation about how the structural parameters in (*), λ, β, and θ, can be identified based on variation in classroom sizes. For ease of discussion and without loss of generality, let us assume for now that all peer effects are positive. Since individuals are excluded from their own peer groups (i.e. exclusive averaging), individuals with attributes above the average essentially have a group of peers whose attributes are below average, which implies a perfect negative correlation between individual attributes and their mean peer attributes. If endogenous and exogenous peer effects exist, it is clear from equation (1) in the main text that such negative correlation reduces the dispersion in outcomes (over the case where peer effects are zero). However, the degree of the smoothing is not constant across demographic groups of different sizes. The smoothing effect is larger in smaller groups since exclusive averaging doesn't greatly effect peer means when group size is large. This is seen in equation (A.3) where the marginal effect of $\mathbf{z}_{ic} - \bar{\mathbf{z}}_c^{(1,h_{1,ic})}$

is reduced by θ (i.e. exogenous peer effects) *linearly* and by λ (i.e. endogenous peer effects) *nonlinearly*, and the reduction decreases as demographic group size, $n_c(1, r)$, increases.

This implies that, as group size varies across classrooms, so does the impact of each structural parameter on the outcome in the model. Thus, these variations can be used to separately identify the parameters. Note that the variation in group sizes in our model is richer than that in the model of Lee (2007). The sizes of multiple demographic groups within a classroom vary across classrooms, which allows us to use the variation in $n_c(k, r)$ for $k = 1, ..., K$ and $r = 1, ..., R_k$, not just n_c.[3]

Another important feature of our model is the additional term (**). Interestingly, the social effects (λ and θ) in this term no longer reduce the dispersion in outcomes, but amplify it. When there are disparities in attributes across demographic groups, the social effects exaggerate the gap. This is intuitive since the social effects work primarily within demographic groups in our model. This gives us an important implication that peer interactions may equalize students outcomes by improving the performance of low ability students more than that of high ability students, but it may also increase the gap across different peer groups when students interact exclusively within their peer groups. Therefore, in this setting, if there is a heterogeneous social shock that affects demographic groups disproportionately (or if there is a pre-existing gap in performance between groups), peer interaction will exaggerate the gap.[4] This suggests that if a gap exists in average performance for members of different groups within a partition, and students interact within their group, mere exposure in the classroom to other groups is not be sufficient to close performance gaps.[5]

Next, we show how variations in partition structures and groups sizes together yield identification in our model in general, for which we use the identification results from Bramoullé et al. (2009). From an instrumental variable perspective, identification of the within-transformed

[3]In equation (A.3), note that $\mathbf{d}_{ic}$ is a vector of partition dummies, so $\mathbf{d}_{ic} = \bar{\mathbf{d}}_c^{(1,h_{1,ic})}$ when $K = 1$. Equation (A.3) implies the coefficients on the dummies can be identified once λ is identified.

[4]For example, if the current pandemic has disproportionate effects on student outcomes across ethnic groups, peer interaction within ethnic groups will amplifies the disparity.

[5]A simple calculation using equation (A.3) reveals that a group's unit deviation from its classroom average in academic performance is increased by $\frac{\lambda}{1-\lambda}$ percent due to peer effects.

equation (A.1) reduces to whether we can find valid instruments for the endogenous regressors $\{\mathbf{W}_{k,c}\mathbf{y}_c^*\}_{k=1}^K$. Note from the reduced form equation (A.2) that $\mathbf{W}_{k,c}\mathbf{y}_c^*$ can be written as

$$\mathbf{W}_{k,c}\mathbf{y}_c^* = \left(\mathbf{W}_{k,c} + \mathbf{W}_{k,c}\sum_{k=1}^K \lambda_k \mathbf{W}_{k,c} + \mathbf{W}_{k,c}\left(\sum_{k=1}^K \lambda_k \mathbf{W}_{k,c}\right)^2 + ...\right)\left[\mathbf{X}_c^*\beta + \sum_{k=1}^K \mathbf{W}_{k,c}\mathbf{Z}_c^*\theta_k + \mathbf{u}_c^*\right] \tag{A.4}$$

For the instrumental variable approach to work, the deterministic parts of $\{\mathbf{W}_{k,c}\mathbf{y}_c^*\}_{k=1}^K$, which simply exclude the error term, $\mathbf{u}_c^*$, from (A.4), should not be perfectly collinear with the exogenous regressors in the model, $\mathbf{X}_c^*$ and $\{\mathbf{W}_{k,c}\mathbf{Z}_c^*\}_{k=1}^K$. Otherwise, identifying instruments for the endogenous terms that satisfy the exogeneity, relevance, and rank conditions don't exist in the system. Therefore, identification of (A.1) requires the deterministic parts of $\{\mathbf{W}_{k,c}\mathbf{y}_c^*\}_{k=1}^K$ to be linearly independent of $\mathbf{X}_c^*$ and $\{\mathbf{W}_{k,c}\mathbf{Z}_c^*\}_{k=1}^K$ for some c. If it is satisfied, the elements of the deterministic part of $\mathbf{W}_{k,c}\mathbf{y}_c^*$ that are independent of the existing regressors can be used as valid instruments for them. This is the insight of Bramoullé et al. (2009) and others, and we show below that the condition is in general satisfied in our case if there is sufficient variation in partition structures and group sizes.

First, observe that $\mathbf{W}_{k,c}\mathbf{W}_{l,c}\mathbf{Z}_c^*$ for $k, l = 1, ..., K$, are elements of the deterministic parts of $\{\mathbf{W}_{k,c}\mathbf{y}_c^*\}_{k=1}^K$, and whose ith element, $[\mathbf{W}_{k,c}\mathbf{W}_{l,c}\mathbf{Z}_c^*]_i$, is given by:

$$\begin{aligned}
&= \frac{1}{n_c(k,h_{k,ic})-1}\sum_{\substack{q\in\mathcal{S}_c(k,h_{k,ic})\\ q\neq i}}\left(\frac{1}{n_c(l,h_{l,qc})-1}\sum_{\substack{j\in\mathcal{S}_c(l,h_{l,qc})\\ j\neq q}}(\mathbf{z}_{jc}-\bar{\mathbf{z}}_c)\right)\\
&= \frac{1}{n_c(k,h_{k,ic})-1}\left\{\sum_{r=1}^{R_k}\left(\frac{n_c([k,h_{k,ic}]\cap[l,r])}{n_c(l,r)-1}\sum_{j\in\mathcal{S}_c(l,r)}(\mathbf{z}_{jc}-\bar{\mathbf{z}}_c)\right)-\frac{1}{n_c(l,h_{l,ic})-1}(\mathbf{z}_{ic}-\bar{\mathbf{z}}_c)\right.\\
&\qquad\left.-\frac{1}{n_c(l,h_{l,ic})-1}\sum_{\substack{j\in\mathcal{S}_c(l,h_{l,ic})\\ j\neq i}}(\mathbf{z}_{jc}-\bar{\mathbf{z}}_c)-\sum_{\substack{j\in\mathcal{S}_c(k,h_{k,ic})\\ j\neq i}}\frac{1}{n_c(l,h_{l,jc})-1}(\mathbf{z}_{jc}-\bar{\mathbf{z}}_c)\right\},
\end{aligned}$$

where $n_c([k,s]\cap[l,r])$ denotes $|\mathcal{S}_c(k,s)\cap\mathcal{S}_c(l,r)|$ (i.e. the number of students in the intersection of group s on partition k and group r on partition l). The first equality implies that the term includes the attributes of student i's peers' peers where direct peers are based on the k^{th} partition and indirect peers are based on the l^{th} partition, and the peer sums are divided by (one less than) the sizes of the demographic groups associated with them. Note that $[\mathbf{Z}_c^*]_i = \mathbf{z}_{ic} - \bar{\mathbf{z}}_c$, $[\mathbf{W}_{l,c}\mathbf{Z}_c^*]_i = \frac{1}{n_c(l,h_{l,ic})-1}\sum_{\substack{j\in\mathcal{S}_c(l,h_{l,ic})\\ j\neq i}}(\mathbf{z}_{jc} - \bar{\mathbf{z}}_c)$, and $[\mathbf{W}_{k,c}\mathbf{Z}_c^*]_i$ is similarly given. Then, from the second equality, it is clear that $\mathbf{W}_{k,c}\mathbf{W}_{l,c}\mathbf{Z}_c^*$ for $k, l = 1, ..., K$ includes terms similar to $\mathbf{Z}_c^*$, $\mathbf{W}_{l,c}\mathbf{Z}_c^*$, and $\mathbf{W}_{k,c}\mathbf{Z}_c^*$, but they won't be linearly dependent as long as there is variation in how each partition intersects with the others, i.e. $\mathcal{S}_c(k,s)\cap\mathcal{S}_c(l,r)$ or variation in group sizes, i.e. $n_c(k,r)$.[6] Since the deterministic parts of $\{\mathbf{W}_{k,c}\mathbf{y}_c^*\}_{k=1}^K$ include $\mathbf{W}_{k,c}\mathbf{W}_{l,c}\mathbf{Z}_c^*$, this ultimately implies that identification of equation (A.1) is possible when such variation exists. When the model is identified, it is clear that $\mathbf{W}_{k,c}\mathbf{W}_{l,c}\mathbf{Z}_c^*$ for $l = 1, ..., K$ can be used as identifying instruments for $\mathbf{W}_{k,c}\mathbf{y}_c^*$.[7]

The role of variations in group sizes and partition structures in the identification of the model can be understood intuitively as follows. Consider the case with two partitions, gender and ethnicity. If peer effects work through both gender and ethnic lines, peer effects within girl and boy groups will first reduce the dispersion in outcomes within the two gender groups, but increase the gap between the two. However, peer interaction within groups in the ethnic partition (ethnic groups) may play against these effects, lessening the gender gap within the ethnic groups but increasing dispersion in outcomes within gender groups by enlarging racial gap in outcomes there. The size of all of these smoothing and dispersing effects and how one effect interplays with the others depend on the sizes of groups and how groups of one partition intersect with groups of the others.

[6] Using the second equality, one can easily see that the terms are perfectly linearly dependent if some partitions completely overlap (e.g. all boys are race 1 and all girls are race 2, so $\mathbf{W}_{g,c} = \mathbf{W}_{e,c}$ for all c).

[7] One can easily verify that these instruments satisfy the exogeneity, relevance, and rank conditions for valid instruments. More formal sufficient identification condition for our model in this regard may be derived using the identification results for high-order spatial autoregressive models in Lee and Liu (2010).

Appendix B Simulation Experiments

We conduct two experiments to examine 1) if the NYC elementary school data contains enough variations in the sizes of demographic groups across classrooms to consistently estimate model parameters, and 2) if our model can capture the determinants of friendship when friendship is formed based on shared demographic characteristics. Specifically, in the first experiment, we simulate datasets using the proposed model and the empirical moments of the sizes of groups in the NYC school data, and investigate if the model parameters are consistently estimated from the proposed maximum likelihood estimation. The second experiment applies the proposed model to the datasets generated from a standard peer effects model with a single friendship network where the peer network is formed based on some shared characteristics. We examine if the proposed model can identify the determinants of the friendship network in such setting.

For the first experiment, we create C classrooms with classroom sizes randomly drawn from the empirical distribution of class sizes in the NYC data. Then, we create three partition networks by randomly assigning a gender, ethnicity and bus ridership to each student in a classroom. The three networks represent three distinctive network types in terms of the number of related groups and the level of variation in their sizes. First, the gender network ($\mathbf{W}_{1,c}$) includes only two groups, so the size of each gender group in a classroom tends to be large,[8] and it may not vary much across classrooms in part due to the common school practice that roughly matches the proportions of male and female students. This implies identification of gender peer effects may be challenging. Second, the ethnicity network ($\mathbf{W}_{2,c}$), generated based on the empirical distributions of the four largest ethnic groups in the NYC data, contains a moderate number of groups, and their size variations are relatively large in Table 1 in the main text. So, this type of network may not pose the identification difficulty that the gender network may have. The last network, bus rider network ($\mathbf{W}_{3,c}$) is very sparse

[8] As discussed in Section 4.2, endogenous and exogenous peer effects may be weakly identified when the average group size is large relative to the number of groups.

as only about 10% of students are bus riders and there may be several bus routes that further partition the riders.[9] It is well known that this type of network sparsity or exclusion restrictions help identification of peer effects, which is well born out in our simulation results. Regressors include dummies created from the gender, ethnicity and bus ridership variables, a continuous variable, $z_{ic} \sim N(0,1)$, and its interactions with weights matrices for exogenous peer effects. The errors are drawn independently and identically from $N(0,1)$. With all variables generated, the outcome variables are generated from the reduced form of equation (1) in the main text for each classroom. This exercise includes thirteen parameters in total: three endogenous peer effects (λ), three exogenous peer effects (θ), and six coefficients for the rest of the regressors and the variance of the error (β and σ^2). We set $\lambda_1 = \lambda_2 = \lambda_3 = 0.1$, and the rest of the parameter values to be one. We simulate 1,000 times for each $C = \{50, 100, 500\}$.

Table B.1: Experiment 1

		$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\beta}_1$	$\hat{\sigma}^2$
C=50	Bias	0.025	0.008	0.003	-0.043	-0.017	0.001	-0.026	0.012
	SD	0.076	0.042	0.048	0.212	0.111	0.198	0.105	0.040
C=100	Bias	0.012	0.004	0.001	-0.025	-0.006	0.002	-0.017	0.005
	SD	0.056	0.029	0.030	0.149	0.075	0.091	0.076	0.028
C=500	Bias	0.001	0.000	0.000	-0.001	0.000	0.001	-0.002	0.000
	SD	0.016	0.009	0.009	0.045	0.024	0.029	0.023	0.009

Table B.1 reports the bias and standard deviation (SD) of the estimates over 1,000 replications for each case.[10] Overall, both bias and variance of the estimates decrease fast as C increases, and the bias is considerably smaller than the standard deviation, confirming that the empirical distributions of the sizes of groups in the NYC data provide sufficient variations for the consistent estimation of the model parameters. Particularly, the biases and variances of $\hat{\lambda}_1$ and $\hat{\theta}_1$ (i.e. gender- endogenous and exogenous peer effects) are always larger than that of the others peer effects estimators as we expected before.

[9] As mentioned in the main text, when constructing bus route or stop network for our empirical application, those who are not bus riders are ignored, that is, those are isolated from these networks.

[10] β_1 is coefficient on z_{ic} and the results on β_i for $i = 2, ..., 6$ are omitted to save space.

For the second experiment, we simulate a friendship network, $\mathbf{W}_c$, for each classroom using a network formation model frequently used in the literature (e.g. Goldsmith-Pinkham and Imbens, 2013) such that $[\mathbf{W}_c^*]_{ij} = w_{ijc}^*$ where $w_{ijc}^* = \mathbf{1}\left(U_i^c(j) > 0\right) \times \mathbf{1}\left(U_j^c(i) > 0\right)$ for $i \neq j$ and $w_{ijc}^* = 0$ otherwise. $U_i^c(j)$ is the utility for i of forming a link with j and is specified as $U_i^c(j) = \alpha_G \mathbf{1}\left(G_{ic} = G_{jc}\right) + \alpha_E \mathbf{1}\left(E_{ic} = E_{jc}\right) + \epsilon_{ijc}$ and $U_{jc}(i)$ similarly defined.[11] G_{ic} and E_{ic} are the gender and ethnicity variables simulated in the first experiment, and $\epsilon_{ijc} \sim$ iid $N(0,1)$. Then, $\mathbf{W}_c$ is obtained by row-normalizing $\mathbf{W}_c^*$. The network formation model implies that there is a higher chance for two students to form a link if they share the same gender or ethnicity. After including the same set of regressors used in the first experiment, the outcome variable is generated from the reduced form of equation (1) in the main text where $\lambda \mathbf{W}_c$ now substitutes for $\sum_{k=1}^{K} \lambda_k \mathbf{W}_{k,c}$. Then, we apply our partitioned peer effect model with three networks as in the first experiment to these data to investigate if the peer effects estimates deliver any information about the determinants of friendship formation.[12] We set $C = 1{,}000$, $\lambda = 0.3$, and simulate 1,000 times for each $(\alpha_G, \alpha_E) \in \{(1,1), (0.5,1), (1,0.5), (0.1,0.1)\}$ where we vary the parameter values to change the relative importance of the two factors in network formation.

Table B.2: Experiment 2

(α_G, α_E)		$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\lambda}_3$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\beta}_1$	$\hat{\sigma}^2$
(1,1)	Mean	0.066	0.034	-0.003	0.221	0.108	0.003	0.983	1.081
	SD	0.019	0.010	0.010	0.037	0.020	0.031	0.006	0.010
(1,0.5)	Mean	0.071	0.018	0.002	0.241	0.056	-0.007	0.987	1.095
	SD	0.018	0.010	0.011	0.037	0.019	0.030	0.006	0.010
(0.5,1)	Mean	0.032	0.041	-0.002	0.132	0.129	-0.004	0.990	1.116
	SD	0.019	0.010	0.011	0.038	0.019	0.030	0.006	0.010
(0.1,0.1)	Mean	0.011	0.010	-0.003	0.051	0.015	-0.004	0.999	1.225
	SD	0.018	0.010	0.011	0.036	0.019	0.029	0.006	0.011

Table B.2 reports the empirical mean and SD of the estimates over 1,000 simulation draws for each case. Overall, the relative importance of the factors in friendship formation reflects

[11] For simplicity, we assume that both students need to agree to form a link.

[12] Note that the bus ridership variable is not included in the friendship formation model, but the bus ridership network is included when estimating the outcome model.

well on the relative sizes of $\hat{\lambda}_k$ and $\hat{\theta}_k$. In all cases, gender and ethnicity network effects are estimated to be larger than the bus rider network effect on average, which captures the fact that gender and ethnicity are the determinants of the friendship network, but the bus ridership is not. Also, gender and ethnicity network effects are estimated as small when α_G or α_E are small, and vice versa. Proposing a formal test for determinants of network formation based on peer effects estimates from our model in this regard is beyond the scope of this paper, but it is clear that the peer effect estimates of our model may provide important insights into the channels through which students form friendships.[13]

[13]The estimates may better reflect the data generating process of friendship formation if the magnitudes of friendship links are allowed to differ by similarity between students.

Appendix C Robustness Checks

C.1 Randomness of classroom assignment

One key assumption for the causal interpretation of educational peer effect models is that classroom assignment is random (e.g., De Giorgi et al., 2010). While we cannot prove this random assignment, we provide evidence to support this claim for this sample. For each homophilous peer network we include in our model, we would like to show that classroom assignment is not a function of the associated variable. To do this, we consider a series of multinomial logit models of the form:[14]

$$P\left(Class_i = c | x_{ikr}\right) = e\left(\alpha_{krc} + \beta_{krc} x_{ikr}\right) / \sum_{j=1}^{C} e\left(\alpha_{krj} + \beta_{krj} x_{ikr}\right)$$

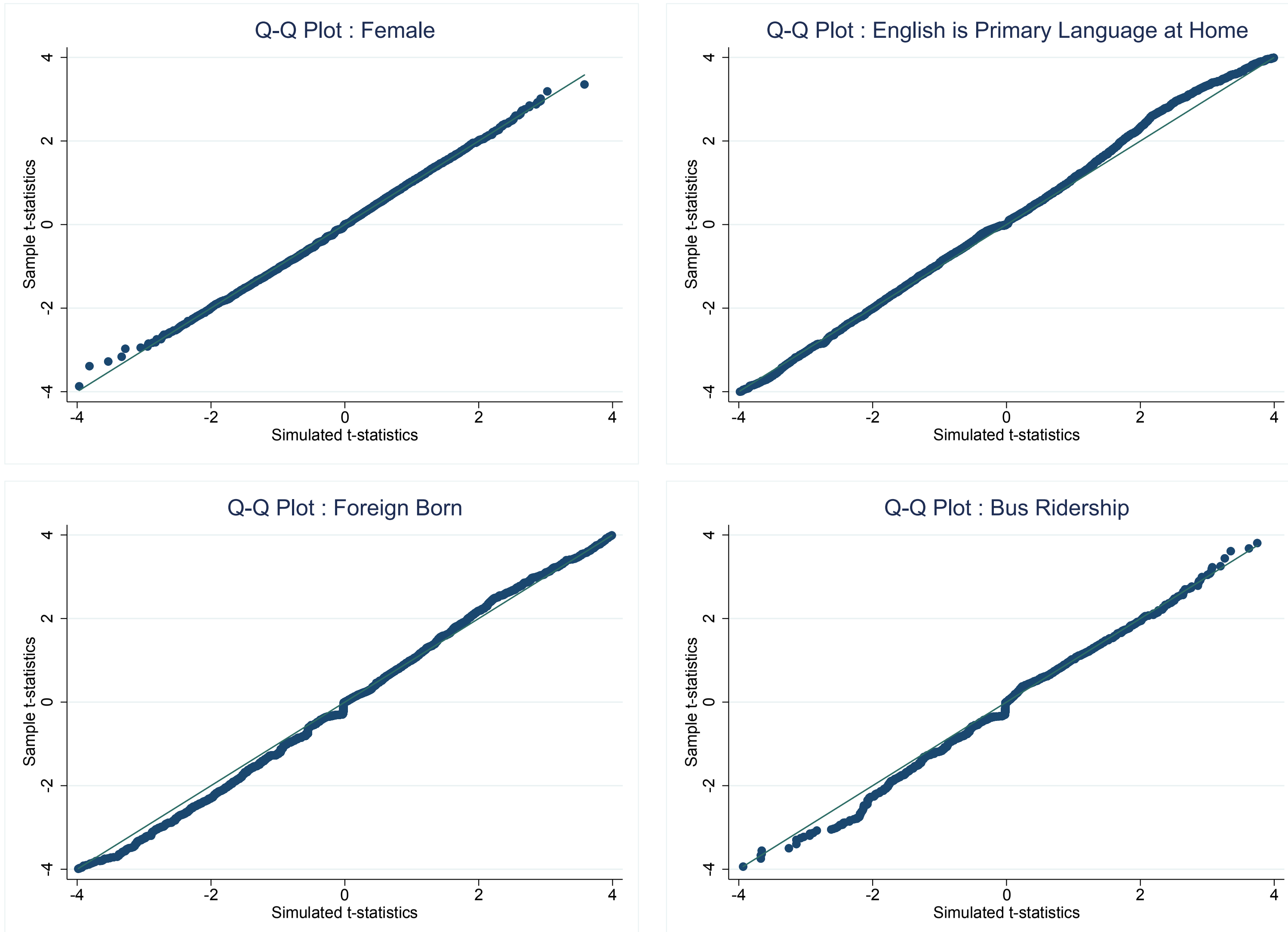
We limit the sample for each regression to a single grade (g) in a single school (s) in a single year (t), which includes 4.44 classrooms on average, and repeat regressions until we cover the whole sample.[15] $Class_i$ is a categorical variable for assignment of student i to one of these classrooms, and x_{ikr} is a binary indicator for group r based on demographic characteristic k of student i that we use as a partition variable in our analysis. For example, if k is the ethnicity network and r is the Hispanic group, x_{ikr} is an indicator of whether student i is Hispanic. We convert demographic indicators with many sparse categories (i.e. 'language spoken at home' and 'bus route') to binary indicators (i.e., 'English speaker' and 'bus rider,' respectively). The result is a series of estimates ($\tilde{\beta}_{krc}$) and t-statistics, capturing the marginal effect and significance of the relevant demographic variable (x_{ikr}) on class assignment. Then, for each school-grade-year we randomly re-assign the students to the classrooms using a uniform distribution, and re-run the regressions. Using a quantile-quantile plot of the simulated and actual t-statistics, we compare the distributions of t-statistics. Each dot in the panels of

[14]The variables and coefifcients in the equation are subscripted with gst, but it is suppressed for notational simplicity.

[15]We exclude school-grades for which there is only one classroom available from this process, as there is no assignment decision to be made.

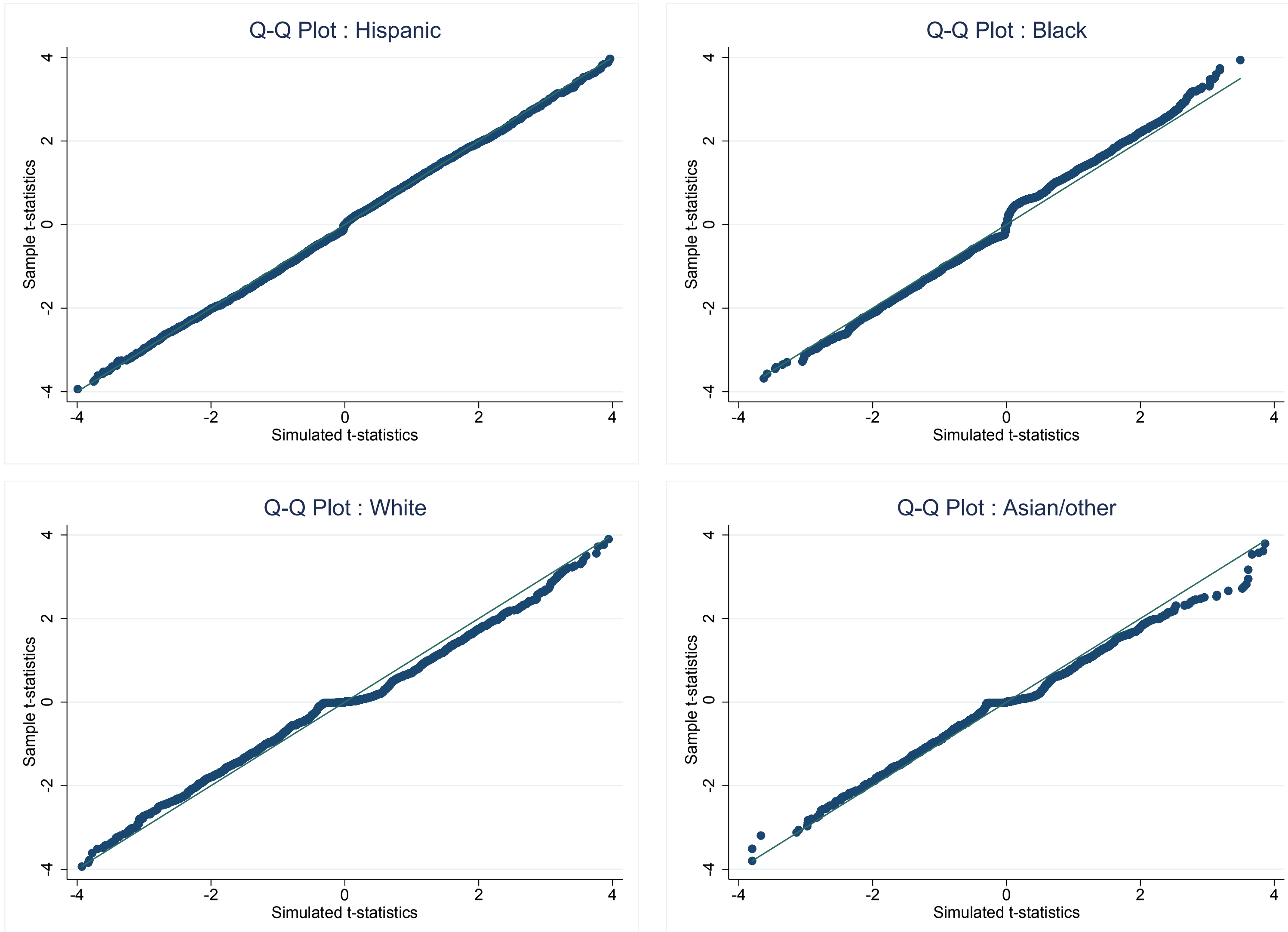
Figures C.1 and C.2 represents a pair of t-statistics (one actual, the other simulated) for each grade-school-year. Each panel represent a single demographic indicator and contains a line with a slope of one (a 45 degree line), so deviations in the dot patterns from this line represent a difference in the significance of the non-randomized and randomized regressions, and (perhaps) a departure from random assignment to classrooms. Figure C.1 contains the plots for female, 'English speaker,' 'foreign born' and 'bus rider,' while figure C.2 contains plots for Hispanic, black, white and Asian. All plots appear to support random assignment of students to classrooms.

Figure C.1: Quantile-Quantile Plots for Student Characteristics



We run a series multinomial logits to estimate the importance of gender, whether English is spoken at home, whether a student is foreign born, or whether bus ridership is important for class assignment. In each graph, we plots the t-statistics from these regressions against t-statisitics from a similar exercise in which we randomly assign students to classrooms. Thus we are plotting these two distributions against one another, and if the distributions are the same, we expect a straight line of slope one. We argue that these provide evidence that class assignment is consistent with a random process.

Figure C.2: Quantile-Quantile Plots for Ethnicity



We run a series multinomial logits to estimate the importance of each ethnic group in class assignment. In each pair of graphs, the left plots the t-statistics from these against the t-statisitics from a similar exercise in which we randomly assign students to classrooms. Thus we are plotting these two distributions against one another, and if the distributions are the same, we should expect a straight line of slope one. We argue that these provide evidence that class assignment is consistent with a random process.

C.2 Past Performance Network

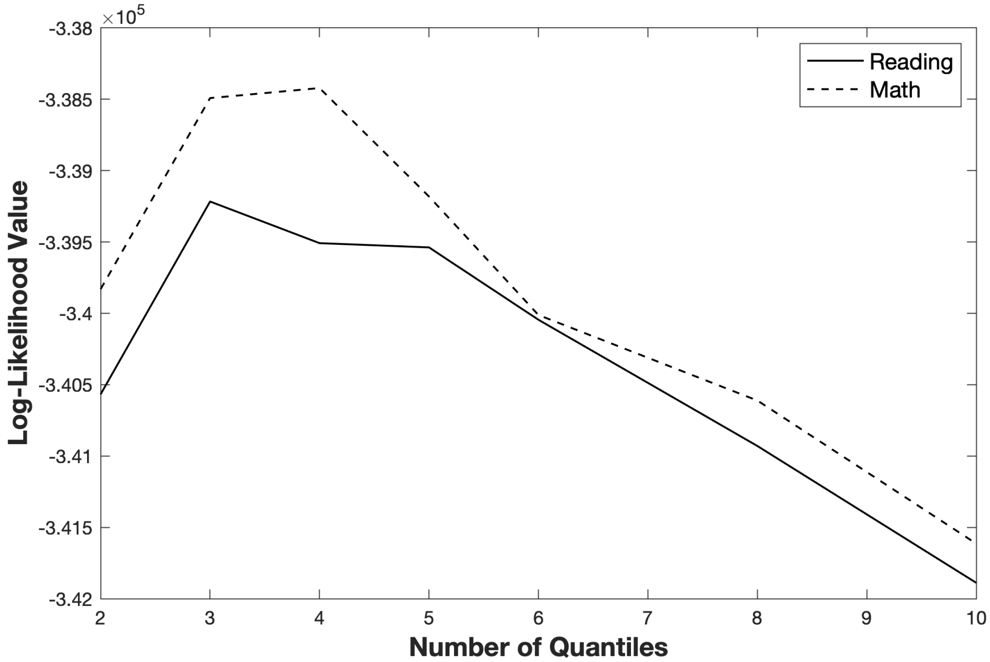
One concern is that students may interact along performance levels, such that high performing students interact more with other high performing students (e.g., Carrell et al., 2013). If this is true and these missing links are associated with our demographic networks, our peer effect estimates across demographic partitions (particularly their rankings) will be biased. We test the robustness of our estimates to this possibility (i.e., missing network bias) by including a network based off lag performance quantile. This network will be treated as a control and not as a network of interest. We rank students in each classroom based on their previous year's test scores and then divide them into groups based on quantile. Using these groups, we construct a past performance network as described in Section 4.1.

One empirical issue is to determine the number of quantiles to form the performance groups. Figure C.3 shows the log-likelihood values of the models with different number of quantiles for the past performance network. We can see that in case of reading, using three quantiles results in the largest likelihood value, while four quantiles results in the largest value for math. Since we consider the past performance network as a control, we choose the number of quantiles that most improves model fit. These models are essentially equivalent to the model (2) and (4) in Table 3 except that they include the addition of a past performance network.

The results are reported in Table C.1. Notice that networks with the largest estimates in the baseline model remain among the most influential networks here. For math, the Ethnicity network is the largest in both the baseline and this model. For reading, the top network is Gender in both models. For math, the Language and then Gender networks are next most influential in the baseline model, although they are superseded here by the Bus networks and Country of birth. The Bus Stop and Bus Route networks are of opposite sign, suggesting there may be significant overlap between the groups.

For reading, Ethnicity remains the second most important network. Language was the third most influential network in the baseline model, but here it is superseded by Country of

Figure C.3: Log-Likelihood Value by the Number of Quantiles for the Past Performance Network



birth, which now shows a similar estimate (0.057) to Language (0.054). One key difference between this model and the baseline model is that all of the networks now appear important. Rather than the lag performance network removing all variation, it has helped to highlight which networks are important, and we see that the primary network for math is Ethnicity, and for reading it is Gender - consistent with our baseline model.

Table C.1: Including a Past Performance Network Control

	Mathematics		Reading	
Peer Effects (λ_k)				
Ethnicity	0.137**	(0.008)	0.093**	(0.004)
Language	0.046**	(0.007)	0.054**	(0.003)
Gender	0.063**	(0.012)	0.215**	(0.007)
Zip Code	0.042**	(0.007)	0.050**	(0.003)
Bus Route	-0.067**	(0.009)	0.015**	(0.003)
Bus Stop	0.085**	(0.010)	-0.002	(0.003)
Country	0.075**	(0.012)	0.057**	(0.004)
Lag Score Qtile	0.452**	(0.005)	0.468**	(0.004)
Observations (n)	55,598		55,221	

C.3 Alternative Specifications

Tables C.2-C.4 modify the specification in Table 3 in the main text to see if our main results are sensitive to different specifications.[16] The general finding of this exercise is that the relative rankings of the peer effects tend to be robust, while the magnitudes vary slightly. For this reason, the baseline model reported in Section 5 and Table 3 in the main text is our preferred specification. The baseline model provides a clear understanding of the relative importance of all partitions of interest, and the impact of the characteristics these partitions represent on elementary school education test scores.

Table C.2 contains the same model as in Table 3 in the main text (equation 2), but removes the bus networks. The bus network tends to be sparser then the other networks, and we want to be sure that this sparseness is not corrupting the other estimated peer effects. For mathematics test scores, both the ranking and estimates of the peer effects are relatively unchanged without the bus networks. Only Zip Code increases in both models (although it is never significant in the reading models), and Ethnicity increases for reading. But these changes are slight and not meaningful. The increase in Zip Code estimate is consistent with

[16] Given that Table 3 in the main text showed little impact based on our definition of Neighborhood, we report estimates using zip codes as they are substantially less computationally intensive.

the idea that the bus networks are small geographic networks and removing them from the model slightly increases the influence of the Zip Code network. Like most large cities, New York is geographically segregated by ethnicity and income, so it may be surprising that Ethnicity is not more affected by the removal of the bus networks.[17]

Next, we further explore the primary Language spoken at home and Country of birth networks. Nearly 65% of our sample speaks English at home, and over 90% of our sample was born in the US. We remove connections due to these networks in order to test whether their inclusion is driving our estimates for these networks. Table C.3 shows estimates of the baseline model with connections for English language and birth in the US removed. We see that the importance of Ethnicity declines in both models and the network strength for both primary Language spoken at home and Country of birth increases. For reading, the increase in the Country of birth network is enough that it becomes statistically significant (at the 5% level), which was not in the baseline model. In mathematics, primary Language was already the second most important network, and we see the estimate increase while the Ethnicity estimate decreases. This suggests that part of what makes the Ethnicity network so important is shared cultural elements such as language.

Because gender and ethnicity are the most commonly investigated demographic determinants of educational outcomes in the literature, we run the model with just these two networks for comparison. Table C.4 contains the results with only Gender and Ethnicity peer networks. In these models Ethnicity peers are important for mathematics scores (0.247) with Gender being only mildly important (0.055), but not for reading scores where Gender (0.150) and Ethnicity (0.129) are much more similar in importance. This is consistent with the findings in Table 3 in the main text, but we notice that the inclusion of other networks such as primary Language spoken at home reduces the effect of Ethnicity, but not Gender. This suggests that Ethnicity is a proxy for some of these other important networks, but

[17] It is important to note that only about 25% of the sample is assigned a bus route (compared with 10% for NYC as a whole). So even as we focus more on the bus, these networks still affect only a portion of our sample - whereas all students participate in the other networks.

Gender is orthogonal to them.

Table C.2: No Bus Networks

	Mathematics		Reading	
Peer Effects (λ_k)				
Ethnicity	0.221	(0.008)	0.117	(0.009)
Language	0.097	(0.009)	0.052	(0.009)
Gender	0.054	(0.013)	0.150	(0.012)
Zip Code	0.020	(0.009)	0.008	(0.009)
Country	0.008	(0.016)	0.024	(0.017)
Observations (n)	55,598		55,221	

About 25% of the sample is assigned a bus, but only a subset of these students share a bus route or stop with a classmate. We run the baseline model without the bus networks, as we may be concerned about sparse networks. Models include 20 contextual social effects, constructed using the lag test score, age, age squared, and FRPL status along with each of the seven networks included in the model, as well as neighborhood and classroom fixed effects. Hispanic is the omitted reference group for ethnicity marginal effects.

C.4 Classroom Composition: Differences by Borough

The NYC public school system is a large and diverse place containing five boroughs each with their own characteristics. We split our sample by borough in order to explore whether our estimates change based on context and by how much. Table C.5 report the results for mathematics and reading respectively when we split our sample by borough. For both outcomes, the top three effects are consistent with the full sample in most splits, but there is some significant variation in the estimates that is worth unpacking.

Starting with mathematics, we notice that in all models the Ethnicity network remains the most important, but point estimates range from 0.136 in the Bronx to 0.285 in Queens. There are likely two causes to this, both of which are tied to the ethnic composition of the boroughs. The first is that the relevance of the Ethnicity network may vary between ethnic groups. Second, whether a student is participating in a large or small network may affect the importance of the network. We can think of this second mechanism as a question

Table C.3: Alternate Language and Country Networks

	Mathematics		Reading	
Peer Effects (λ_k)				
Ethnicity	0.208	(0.008)	0.115	(0.009)
Language	0.125	(0.009)	0.053	(0.010)
Gender	0.054	(0.013)	0.150	(0.012)
Zip Code	0.015	(0.009)	0.004	(0.009)
Bus Route	0.007	(0.012)	0.028	(0.012)
Bus Stop	0.019	(0.013)	0.007	(0.013)
Country	0.017	(0.016)	0.038	(0.017)
Observations (n)	55,598		55,221	

Models include 28 contextual social effects, constructed using the lag test score, age, age squared, and FRPL status along with each of the seven networks included in the model, as well as neighborhood and classroom fixed effects. Hispanic is the omitted reference group for ethnicity marginal effects.

Table C.4: Gender and Ethnicity

	Mathematics		Reading	
Peer Effects (λ_k)				
Ethnicity	0.247	(0.008)	0.129	(0.009)
Gender	0.055	(0.013)	0.150	(0.012)
Observations (n)	55,598		55,221	

Models include eight contextual social effects, constructed using the lag test score, age, age squared, and FRPL status along with each of the seven networks included in the model, as well as neighborhood and classroom fixed effects. Hispanic is the omitted reference group for ethnicity marginal effects.

Table C.5: Sample Partitioned by Borough (Mathematics)

Panel A: Mathematics	Bronx	Brooklyn	Manhattan	Queens	Staten Island
Peer Effects (λ_k)					
Ethnicity	0.136	0.199	0.230	0.285	0.158
	(0.021)	(0.020)	(0.025)	(0.012)	(0.016)
Language	0.081	0.078	-0.002	0.118	0.060
	(0.029)	(0.021)	(0.035)	(0.012)	(0.019)
Gender	0.019	0.081	0.019	0.029	0.096
	(0.035)	(0.032)	(0.044)	(0.021)	(0.023)
Census Tract	0.028	0.010	0.013	0.002	0.049
	(0.022)	(0.019)	(0.022)	(0.014)	(0.019)
Bus Stop	0.107	0.031	-0.013	0.005	0.020
	(0.037)	(0.038)	(0.049)	(0.020)	(0.020)
Bus Route	0.025	-0.011	0.047	0.010	-0.003
	(0.029)	(0.031)	(0.044)	(0.020)	(0.018)
Country	-0.013	0.041	0.065	-0.006	0.007
	(0.042)	(0.034)	(0.056)	(0.021)	(0.040)
Observations (n)	6,385	8,908	4,181	19,794	16,330

Panel B: Reading	Bronx	Brooklyn	Manhattan	Queens	Staten Island
Peer Effects (λ_k)					
Ethnicity	0.080	0.090	0.120	0.153	0.095
	(0.025)	(0.022)	(0.024)	(0.015)	(0.017)
Language	0.012	0.060	-0.030	0.071	0.037
	(0.034)	(0.022)	(0.031)	(0.013)	(0.019)
Gender	0.197	0.107	0.163	0.129	0.171
	(0.033)	(0.031)	(0.035)	(0.021)	(0.022)
Census Tract	0.025	0.004	0.014	-0.013	0.01989
	(0.026)	(0.020)	(0.020)	(0.015)	(0.019)
Bus Stop	0.034	0.019	0.024	-0.022	0.028
	(0.042)	(0.039)	(0.046)	(0.023)	(0.021)
Bus Route	0.044	0.025	0.037	0.043	0.008
	(0.036)	(0.032)	(0.040)	(0.022)	(0.019)
Country	-0.038	-0.036	0.100	0.034	0.069
	(0.052)	(0.037)	(0.052)	(0.024)	(0.042)
Observations (n)	6,367	8,811	4,177	19,559	16,307

Models include 28 contextual social effects, constructed using the lag test score, age, age squared, and FRPL status along with each of the seven networks included in the model, as well as neighborhood and classroom fixed effects. Hispanic is the omitted reference group for ethnicity marginal effects.

of the level of diversity in a classroom, and we discuss this mechanism further in Section 6. The second and third most important networks are usually primary Language spoken at home and Gender, although there is more variation than there was in Ethnicity, and our estimates may be less precise due to smaller samples in these splits. Turning to the results of each borough, Manhattan is the smallest borough, and only the ethnicity network is significant for mathematics. In the other boroughs, gender is the second most important network for Brooklyn and Staten Island, but it is insignificant for Queens and the Bronx. Language is the second most important network in Queens, and the third most important in the Bronx, Brooklyn, and Staten Island. The estimate in Queens is large (0.118), over double the estimate for the full model. Queens is also the only borough in which English is not the primary Language for most students (48%), although it is still the modal language. In the Bronx, it is notable that the second most important network is the Bus Stop, and Neighborhood is significant (at least at the 10% level) for the Bronx, Brooklyn, and Staten Island. These Neighborhood estimates may suggest differences in the role Neighborhood plays in different boroughs.

For reading, we notice that Gender is the most important network in all but one borough, and Ethnicity is the second most important network for all but one borough. The exception is Queens in which Gender and Ethnicity are swapped, but still remain the top two networks. This again reflects our results in the full model, but there is interesting variation in estimates to note. First, Gender varies greatly from the Bronx (0.197) to Brooklyn (0.107). With the exception of Brooklyn, we see a pattern that as the strength of the Ethnicity network increases, the Gender network decreases. This could suggest that the Ethnicity and Gender networks act as substitutes. A similar pattern occurred for mathematics as well, further strengthening this possibility. We will look more at this in the next section, where we dive further into the Ethnicity network. We should also note the results of the bus networks. For Queens, the Bus Route is the fourth most important network, and it has a sizable effect (0.043). The Bronx, Brooklyn, and Manhattan all have estimates of similar magnitude, but the results are noisy and not statistically significant. This may be due to low power, as these are the smallest boroughs. It is striking that Staten Island has such a low (and insignificant) estimate - even as bus ridership is most common here. This may indicate other factors around how these networks operate. In Staten Island, bus routes may be less tied to neighborhood than the rest of NYC, as the distances buses travel is further in Staten Island than in other boroughs. This may point to the relevant mechanism not being time spent on the bus, but rather the students who live most nearby - a micro neighborhood.

Appendix D Estimation procedure

Our empirical model is given by

$$\mathbf{y}_{ct} = \sum_{k=1}^{K} \lambda_k \mathbf{W}_{k,ct}\mathbf{y}_{ct} + \mathbf{X}_{ct}\beta + \sum_{k=1}^{K} \mathbf{W}_{k,ct}\mathbf{Z}_{ct}\theta_k + \mathbf{y}_{c,t-1}\gamma + \delta_{ct} \cdot \iota_{n_{ct}} + \mathbf{u}_{ct} \tag{D.1}$$

In order to apply a quasi-MLE, we assume that each element in $\mathbf{u}_{ct}$ is $iid(0, \sigma^2)$. To remove the time varying classroom fixed effects, following Lee et al. (2010), we consider an transformation that eliminates the classroom fixed effects while maintaining interdependency between the errors. Let the orthonormal matrix of the with transformation matrix $\mathbf{Q}_{ct} = \mathbf{I}_{n_{ct}} - \iota_{n_{ct}}\iota_{n_{ct}}^{'}/n_{ct}$ be $[\mathbf{P}_{ct}, \iota_{n_{ct}}/\sqrt{n_{ct}}]$. The columns in $\mathbf{P}_{ct}$ are eigenvectors of $\mathbf{Q}_{ct}$ corresponding to the eigenvalue one, such that $\mathbf{P}_{ct}^{'}\iota_{n_{ct}} = \mathbf{0}$, $\mathbf{P}_{ct}^{'}\mathbf{P}_{ct} = \mathbf{I}_{n_{ct}-1}$ and $\mathbf{P}_{ct}\mathbf{P}_{ct}^{'} = \mathbf{Q}_{ct}$. Then, premultiplying equation (D.1) by $\mathbf{P}_{ct}^{'}$ yields

$$\overline{\mathbf{y}}_{ct} = \sum_{k=1}^{K} \lambda_k \overline{\mathbf{W}}_{k,ct}\overline{\mathbf{y}}_{ct} + \overline{\mathbf{X}}_{ct}\beta + \sum_{k=1}^{K} \overline{\mathbf{W}}_{k,ct}\overline{\mathbf{Z}}_{ct}\theta_k + \overline{\mathbf{y}}_{c,t-1}\gamma + \overline{\mathbf{u}}_{ct} \tag{D.2}$$

where $\overline{\mathbf{y}}_{ct} = \mathbf{P}_{ct}^{'}\mathbf{y}_{ct}$ and $\overline{\mathbf{X}}_{ct}, \overline{\mathbf{y}}_{c,t-1}, \overline{\mathbf{Z}}_{ct}$, and $\overline{\mathbf{u}}_{ct}$ are similarly defined, and $\overline{\mathbf{W}}_{k,ct} = \mathbf{P}_{ct}^{'}\mathbf{W}_{k,ct}\mathbf{P}_{ct}$ for $k = 1, ..., K$. In equation (D.2), we use the fact $\mathbf{P}_{ct}^{'}\mathbf{W}_{k,ct} = \overline{\mathbf{W}}_{k,ct}\mathbf{P}_{ct}^{'}$.

Note that $\overline{\mathbf{u}}_{ct} \sim (\mathbf{0}, \sigma^2\mathbf{I}_{n_{ct}-1})$. Then, the likelihood function for (D.2) is given by

$$\ln L_{ct}(\mathbf{\Lambda}, \mathbf{B}, \sigma^2) = -\frac{n_{ct}-1}{2}\ln(2\pi\sigma^2) + \ln|\overline{\mathbf{S}}_{ct}(\mathbf{\Lambda})| - \frac{\bar{\epsilon}_{ct}(\psi)^{'}\bar{\epsilon}_{ct}(\psi)}{2\sigma^2} \tag{D.3}$$

where $\mathbf{\Lambda} = (\lambda_1, ..., \lambda_K)^{'}$, $\mathbf{B} = (\beta^{'}, \mathbf{\Theta}^{'}, \gamma^{'})^{'}$ with $\mathbf{\Theta} = (\theta_1^{'}, ..., \theta_K^{'})^{'}$, $\overline{\mathbf{S}}_{ct}(\mathbf{\Lambda}) = \mathbf{I}_{n_{ct}} - \sum_{k=1}^{K} \lambda_k \overline{\mathbf{W}}_{k,ct}$, and $\bar{\epsilon}_{ct}(\psi) = \overline{\mathbf{y}}_{ct} - \sum_{k=1}^{K} \lambda_k \overline{\mathbf{W}}_{k,ct}\overline{\mathbf{y}}_{ct} - \overline{\mathbf{X}}_{ct}\beta - \sum_{k=1}^{K} \overline{\mathbf{W}}_{k,ct}\overline{\mathbf{Z}}_{ct}\theta_k - \overline{\mathbf{y}}_{c,t-1}\gamma$ with $\psi = (\mathbf{\Lambda}, \mathbf{B})$.

If no one is isolated from any of the networks, following Lee et al. (2010), (D.3) can be written without $\mathbf{P}_{st}$ as

$$\ln L_{ct}(\mathbf{\Lambda}, \mathbf{B}, \sigma^2) = -\frac{n_{ct}-1}{2}\ln(2\pi\sigma^2) - \ln\left(1 - \sum_{k=1}^{K}\lambda_k\right) + \ln|\mathbf{S}_{ct}(\mathbf{\Lambda})| - \frac{\epsilon_{ct}(\psi)^{'} \cdot \mathbf{Q}_{ct} \cdot \epsilon_{ct}(\psi)}{2\sigma^2}$$

$$(D.4)$$

where $\mathbf{S}_{ct}(\mathbf{\Lambda}) = \mathbf{I}_{n_{ct}} - \sum_{k=1}^{K} \lambda_k \mathbf{W}_{k,ct}$, and $\epsilon_{ct}(\psi) = \mathbf{y}_{ct} - \sum_{k=1}^{K} \lambda_k \mathbf{W}_{k,ct}\mathbf{y}_{ct} - \mathbf{X}_{ct}\beta - \sum_{k=1}^{K} \mathbf{W}_{k,ct}\mathbf{Z}_{ct}\theta_k - \mathbf{y}_{c,t-1}\gamma$. The parameter space for $\mathbf{\Lambda}$ needs to be restricted to ensure that $|\mathbf{S}_{ct}(\mathbf{\Lambda})|$ and $\left(1 - \sum_{k=1}^{K} \lambda_k\right)$ are strictly positive for all c and t, so that the likelihood is well defined. $|\mathbf{S}_{ct}(\mathbf{\Lambda})|$ will be strictly positive when $\sum_{k=1}^{K} |\lambda_k| < 1$ since our network matrices are row-normalized.

The log-likelihood function for entire sample is simply $\ln L(\mathbf{\Lambda}, \mathbf{B}, \sigma^2) = \sum_{c=1}^{C} \sum_{t=1}^{T} \ln L_{ct}(\mathbf{\Lambda}, \mathbf{B}, \sigma^2)$. To simplify the estimation, we concentrate out $\mathbf{B}$ and σ^2 in (D.4) using the fact that the QMLE of $\mathbf{B}$ and σ^2 given $\mathbf{\Lambda}$ are: $\widehat{\mathbf{B}}(\mathbf{\Lambda}) = \mathbf{\Phi}^{-1}\mathbf{\Psi}(\mathbf{\Lambda})$ where $\mathbf{\Phi} = \sum_{c=1}^{C} \sum_{t=1}^{T} \chi_{ct}' \cdot \mathbf{Q}_{ct} \cdot \chi_{ct}$ and $\mathbf{\Psi}(\mathbf{\Lambda}) = \sum_{c=1}^{C} \sum_{t=1}^{T} \chi_{ct}' \cdot \mathbf{Q}_{ct} \cdot \mu_{ct}(\mathbf{\Lambda})$ with $\chi_{ct} = [\mathbf{X}_{ct}, \mathbf{W}_{1,ct}\mathbf{Z}_{ct}, ..., \mathbf{W}_{K,ct}\mathbf{Z}_{ct}, \mathbf{y}_{c,t-1}]$ and $\mu_{ct}(\mathbf{\Lambda}) = \mathbf{y}_{ct} - \sum_{k=1}^{K} \lambda_k \mathbf{W}_{k,ct}\mathbf{y}_{ct}$, and

$$\widehat{\sigma}^2(\mathbf{\Lambda}) = \frac{\mathbf{\Upsilon}(\mathbf{\Lambda}) - \mathbf{\Psi}(\mathbf{\Lambda})'\mathbf{\Phi}^{-1}\mathbf{\Psi}(\mathbf{\Lambda})}{\sum_{c=1}^{C} \sum_{t=1}^{T} (n_{ct} - 1)}. \tag{D.5}$$

where $\mathbf{\Upsilon}(\mathbf{\Lambda}) = \sum_{c=1}^{C} \sum_{t=1}^{T} \mu_{ct}(\mathbf{\Lambda})' \cdot \mathbf{Q}_{ct} \cdot \mu_{ct}(\mathbf{\Lambda})$.

Then, the concentrated log-likelihood function of $\mathbf{\Lambda}$ is

$$\ln L(\mathbf{\Lambda}) = \sum_{c=1}^{C} \sum_{t=1}^{T} -\frac{n_{ct} - 1}{2}[\ln(2\pi)+1] - \ln\left(1 - \sum_{k=1}^{K} \lambda_k\right) + \ln|\mathbf{S}_{ct}(\mathbf{\Lambda})| - \frac{n_{ct} - 1}{2} \ln \widehat{\sigma}^2(\mathbf{\Lambda}) \tag{D.6}$$

The QMLE of $\mathbf{\Lambda}$, $\widehat{\mathbf{\Lambda}}$, is the maximizer of (D.6), and the QMLE of $\mathbf{B}$ and σ^2 are $\widehat{\mathbf{B}}(\widehat{\mathbf{\Lambda}})$ and $\widehat{\sigma}^2(\widehat{\mathbf{\Lambda}})$, respectively.

References

Abramson, Jenny, Stephanie Cohen, Julia Feldman, Eva Ho, Dr. Dana Kanze, Deborah Kilpatrick, Pam Kostka, Sol Lee, Devorah Ohana, Julian Salisbury, Noga Tal, Jemma Wolfe, and Steffi Wu (2019). *All In: Women in the VC Ecosystem*. Report. Pitchbook and All Raise.

Ananat, Elizabeth, Fu Shihe, and Stephen L. Ross (2018). "Race-specific urban wage premia and the black-white wage gap". In: *Journal of Urban Economics* 108.C, pp. 141–153.

Arcidiacono, Peter and Sean Nicholson (2005). "Peer effects in medical school". In: *Journal of Public Economics* 89.2-3, pp. 327–350.

Bennett, Magdalena and Peter Bergman (2018). *Better Together? Social Networks in Truancy and the Targeting of Treatment*. IZA Discussion Papers 11267. Institute of Labor Economics (IZA).

Billings, Stephen B., David J. Deming, and Stephen L. Ross (Jan. 2019). "Partners in Crime". In: *American Economic Journal: Applied Economics* 11.1, pp. 126–50.

Boucher, Vincent, Yann Bramoullé, Habiba Djebbari, and Bernard Fortin (2014). "Do Peers Affect Student Achievement? Evidence From Canada Using Group Size Variation". In: *Journal of Applied Econometrics* 29.1, pp. 91–109.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of peer effects through social networks". In: *Journal of Econometrics* 150.1, pp. 41–55.

— (2020). "Peer Effects in Networks: A Survey". In: *Annual Review of Economics* 12.1, pp. 603–629.

Burke, Mary A. and Tim R. Sass (2013). "Classroom Peer Effects and Student Achievement". In: *Journal of Labor Economics* 31.1, pp. 51–82.

Carrell, Scott E., Richard L. Fullerton, and James E. West (2009). "Does Your Cohort Matter? Measuring Peer Effects in College Achievement". In: *Journal of Labor Economics* 27.3, pp. 439–464.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West (2013). "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation". In: *Econometrica* 81.3, pp. 855–882.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014). "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates". In: *American Economic Review* 104.9, pp. 2593–2632.

Damm, Anna Piil (2014). "Neighborhood quality and labor market outcomes: Evidence from quasi-random neighborhood assignment of immigrants". In: *Journal of Urban Economics* 79, pp. 139–166.

De Giorgi, Giacomo, Michele Pellizzari, and Silvia Redaelli (2010). "Identification of Social Interactions through Partially Overlapping Peer Groups". In: *American Economic Journal: Applied Economics* 2.2, pp. 241–75.

Epple, Dennis and Richard E. Romano (2011). "Chapter 20 - Peer Effects in Education: A Survey of the Theory and Evidence". In: ed. by Jess Benhabib, Alberto Bisin, and

Matthew O. Jackson. Vol. 1. Handbook of Social Economics. North-Holland, pp. 1053–1163.

Fruehwirth, Jane Cooley (2013). “Identifying peer achievement spillovers: Implications for desegregation and the achievement gap”. In: *Quantitative Economics* 4.1, pp. 85–124.

Goldsmith-Pinkham, Paul and Guido W. Imbens (2013). “Social Networks and the Identification of Peer Effects”. In: *Journal of Business & Economic Statistics* 31.3, pp. 253–264.

Hegde, Deepak and Justin Tumlinson (2014). “Does Social Proximity Enhance Business Partnerships? Theory and Evidence from Ethnicity’s Role in U.S. Venture Capital”. In: *Management Science* 60.9, pp. 2355–2380.

Hong, Sok Chul and Jungmin Lee (2017). “Who is sitting next to you? Peer effects inside the classroom”. In: *Quantitative Economics* 8.1, pp. 239–275.

Horrace, William C., Hyunseok Jung, and Shane Sanders (2020). “Network Competition and Team Chemistry in the NBA”. In: *Journal of Business & Economic Statistics* 0.0, pp. 1–15.

Horrace, William C., Xiaodong Liu, and Eleonora Patacchini (2016). “Endogenous network production functions with selectivity”. In: *Journal of Econometrics* 190.2, pp. 222–232.

Horrace, William C. and Christopher F. Parmeter (2017). “Accounting for Multiplicity in Inference on Economics Journal Rankings”. In: *Southern Economic Journal* 84.1, pp. 337–347.

Hsieh, Chih-Sheng and Xu Lin (2017). “Gender and racial peer effects with endogenous network formation”. In: *Regional Science and Urban Economics* 67, pp. 135–147.

Kane, Thomas J., Daniel F. Mccaffrey, Trey Miller, and Douglas O. Staiger (2013). “Have we identified effective teachers? Validating Measures of Effective Teaching using Random Assignment”. In: *Research Paper. MET Project. Bill & Melinda Gates Foundation*.

Laschever, Ron A. (June 2013). *The Doughboys Network: Social Interactions and the Employment of World War I Veterans*. Tech. rep. URL: `https://urldefense.com/v3/__https://ssrn.com/abstract=1205543__;!!K543PA!fcKYMUGSjbFVjJOwcftfVg3eXdBanoQXMehrwJqb3ONiOM WIA$`.

Lavy, Victor, Olmo Silva, and Felix Weinhardt (2012). “The Good, the Bad, and the Average: Evidence on Ability Peer Effects in Schools”. In: *Journal of Labor Economics* 30.2, pp. 367–414.

Lee, Lung-Fei (2007). “Identification and estimation of econometric models with group interactions, contextual factors and fixed effects”. In: *Journal of Econometrics* 140.2, pp. 333–374.

Lee, Lung-Fei and Xiaodong Liu (2010). “Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances”. In: *Econometric Theory* 26.1, pp. 187–230.

Lee, Lung-Fei, Xiaodong Liu, and Xu Lin (2010). “Specification and Estimation of Social Interaction Models with Network Structures”. In: *Econometrics Journal* 13, pp. 145–176.

Leszczensky, Lars and Sebastian Pink (2019). “What Drives Ethnic Homophily? A Relational Approach on How Ethnic Identification Moderates Preferences for Same-Ethnic Friends”. In: *American Sociological Review* 84.3, pp. 394–419.

Lett, Elle, H. Murdock, Whitney Orji, Jaya Aysola, and Ronnie Sebro (2019). “Trends in Racial/Ethnic Representation Among US Medical Students”. In: *JAMA Network Open* 2, pp. 1–12.

Manski, Charles F. (1993). “Identification of Endogenous Social Effects: The Reflection Problem”. In: *The Review of Economic Studies* 60.3, pp. 531–542.

Mayer, Adalbert and Steven Puller (2008). “The old boy (and girl) network: Social network formation on university campuses”. In: *Journal of Public Economics* 92.1-2, pp. 329–347.

Nakajima, Ryo (2007). “Measuring Peer Effects on Youth Smoking Behaviour”. In: *The Review of Economic Studies* 74.3, pp. 897–935.

Paloyo, Alfredo R. (2020). “Chapter 21 - Peer effects in education: recent empirical evidence”. In: *The Economics of Education (Second Edition)*. Ed. by Steve Bradley and Colin Green. Academic Press, pp. 291–305.

Patacchini, Eleonora, Edoardo Rainone, and Yves Zenou (2017). “Heterogeneous peer effects in education”. In: *Journal of Economic Behavior and Organization* 134, pp. 190–227.

Presler, Jonathan (2020). *You are who you eat with: Evidence on academic peer effects from school lunch lines*. Working Paper.

— (2022). *Who’s Affecting Who? Obesity Peer Effects in NYC Public Schools*. Working Paper.

Renna, Francesco, Irina B. Grafova, and Nidhi Thakur (2008). “The effect of friends on adolescent body weight”. In: *Economics & Human Biology* 6.3, pp. 377–387.

Sacerdote, Bruce (2001). “Peer Effects with Random Assignment: Results for Dartmouth Roommates”. In: *The Quarterly Journal of Economics* 116.2, pp. 681–704.

Soetevent, Adriaan R. and Peter Kooreman (2007). “A discrete-choice model with social interactions: with an application to high school teen behavior”. In: *Journal of Applied Econometrics* 22.3, pp. 599–624.

Theil, Henri (1972). *Statistical Decomposition Analysis*. Amsterdam: North-Holland Publishing Company.

Weinstein, Meryle, Sarah Cordes, Chris Rick, and Amy Ellen Schwartz (2021). *Riding the Yellow School Bus: Equity in bus transportation across districts, schools, and students*. Working Paper.

White, Michael J. (1986). “Segregation and Diversity Measures in Population Distribution”. In: *Population Index* 52.2, pp. 198–221.

Xu, Yilan and Linlin Fan (2018). “Diverse friendship networks and heterogeneous peer effects on adolescent misbehaviors”. In: *Education Economics* 26.3, pp. 233–252.