HCEO WORKING PAPER SERIES

Peers' Parents and Educational Attainment

Bobby Chung

Working Paper

2018-086

11/2018; Revised 3/2020

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

Peers' Parents and Educational Attainment: The Exposure Effect *

Bobby W. Chung †

March 13, 2020

[Accepted Version for Labour Economics]

Abstract

This paper discusses the 'exposure effect' in child development by investigating the extent to which the educational background of peers' parents is related to a child's future college attainment. I analyze the friendship networks of a nationally representative sample of high-school students in the US. To address endogenous friendship formation, I adopt two distinct strategies: a selection correction approach and exploiting within-school cohort variations in parental compositions. I find that peers' academic performance and other observed characteristics, with a rich set of control variables and network fixed effect, do not fully explain the spillover from peers' parents of the same gender. Effects are more prominent for students with a disadvantaged background - less-educated parents, single-mother households, and less caring fathers. Suggestive evidence is provided to support the role model effect as a plausible channel.

JEL Classification: D91, I24, J16, Z13

Keywords: Peer Effect, College Attainment, Childhood Exposure, Contextual Effect

---

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

†School of Labor and Employment Relations, University of Illinois at Urbana-Champaign, 123 LER Building, Champaign, IL 61820. Email: wychung@illinois.edu. Phone: 1-864-633-8236. This article is a modification of my job market paper. I thank my dissertation committee of Howard Bodenhorn, Peter Blair, Tom Lam, and Kevin Tsui for their patience and support. I am grateful for the comments from Kim-Sau Chung, Steven Durlauf, Alex Eble, Paul Goldsmith-Pinkham, Amanda Griffith, Chih-Sheng Hsieh, Elijah Neilson, Benjamin Posmanick, Kenneth Whaley, Dan Wood, and Jian Zou. Special thanks to the seminar participants at Clemson, EALE, UChicago, Wake Forest, HKBU, and the Fourth Conference on Network Science and Economics. All remaining errors are mine.

1 Introduction

Social interaction affects our beliefs and behaviors, constituting the 'exposure effect' in shaping child development. In the literature of economics, parents and peers have been identified as two indisputable sources. In the context of social network where individuals are directly or indirectly connected, peers' parents can also affect the early-life experience of an individual. Compared to the vast literature on parental influence and peer effect, the relationship between peers' parents and their social influences is relatively understudied.

The influence of peers' parents can be non-negligible for a number of reasons. One mechanism is that children learn from their parents to behave in certain ways and affect their peers. Peers' contemporaneous spillover such as academic performance, which is the primary focus of the peer effect literature, shall be considered a close counterpart of this channel. Parents are also the active actors behind school policies that benefit or harm the peers of their children (Walsh, 2008). Other possibilities involve direct interaction or the spread of information among social ties (Granovetter, 1973; Olivetti et al., 2018; Eble and Hu, 2019). Whereas most of the attention in applied social network analysis is on the spillover among peers, this paper shows that the parent of one's peer is another source of social influences.

To measure the influence of peers' parents, I investigate the extent to which average college attainment of peers' parents is related to a child's future college attainment. I analyze the National Longitudinal Survey of Adolescent Health (AddHealth) dataset which covers a nationally representative sample of high-school students in the US. In addition to standard demographic characteristics, the AddHealth dataset contains the detail of friendship networks in a school and the educational background of parents. The estimates from a baseline binary choice model reveal that the college attainment of peers' parents does influence a child's future college attainment. Moreover, the mechanisms are based on the gender of both peers' parents and the child. Whereas the inclusion of peers' observed characteristics (including academic performance), neighborhood characteristics, and network fixed effects negate the effect of peers' parents of the opposite gender, the spillover from peers' parents of the same gender remains robust. The magnitude of the same gender spillover is smaller compared to the influence of stronger social ties. For example, the spillover of peers' paternal education on boys is one-third of the effect of having a father who is a college graduate, whereas the spillover of peers' maternal education on girls is one-sixth of the effect of having a mother who is a college graduate.

To obtain a causal interpretation, an ideal setting is to randomly assign peers of different family background and observe their interaction with peers' parents. However, in an observational study where friendships were formed naturally, one might worry that non-random sorting among students is a confounding factor. The inclusion of network fixed effects in the baseline model is not sufficient because there may still be within-school unobserved heterogeneity related to both the college outcome and friendship decision, which resembles the omitted variable bias. Therefore, I adopt two distinct approaches. The first strategy involves a selection-corrected model in the spirit of the Heckman two-step procedure (Heckman, 1976, 1979; Hsieh and Lee, 2016). I simultaneously estimate network formation and college outcome equation, taking into account friendship sorting based on similarities in unobserved characteristics.[1] This structural approach addresses a common type of friendship sorting called 'homophily'. For example, smarter students tend to study together. However, there exist other types of formation process.[2] To impose less assumptions on the friendship formation process, my second strategy follows a commonly used approach in the peer effect literature to exploit plausibly idiosyncratic variations in parental composition of each cohort (Hoxby, 2000; Bifulco et al., 2011; Olivetti et al., 2018). This strategy regards all grade-mates as peers, as opposed to actual links, and makes use of quasi-random peer group formations. In both approaches, I do find a similar gender-specific pattern for students from a lower-educated family.

[1] The same-gender pattern remains robust to addressing friendship selection, and the correction in upward bias occurs mainly on the effect of peers' academic performance.

[2] See Jackson (2010) for more discussion on network formation. For example, in the Co-author model, the link decision of an agent also depends on the number of connections of other agents (Jackson and Wolinsky, 1996).

I arrive to the same conclusion as the recent finding by Fruehwirth and Gagete-Miranda (2019): the inclusion of peers' academic performance, which is conventionally regarded as the 'peer effect', does not fully explain the spillover of peers' parents.[3] The gender-specific pattern further complicates the explanations behind the residual influence. Olivetti et al. (2018), who also analyze the AddHealth dataset, provided a clue. They find that the work decision of peers' mothers generates as a role model effect and affects the future labor force participation of a girl. I therefore posit a similar logic about my current finding: there exists a role model effect from peers' parents on human capital acquisition. There are several pieces of supporting evidence. First, the same-gender spillover is only significant when neither or only one of the parents graduated from college. Second, students whose father is absent also experience significant spillover from peers' fathers.[4] The estimate is particularly strong for girls. Third, the magnitude of the same gender spillover on males is stronger when one's own father is less caring. Changing from having the least caring to the most caring father completely offsets the same gender spillover from peers' fathers on males. As in the literature, young people look up to outside adult figures when they lack a role model inside family (Gustafson et al., 1992; Bisin and Verdier, 2001). Last, I exploit the unique but under-utilized aspect of the AddHealth dataset – the ranking of friends. I find that the effect concentrates on the parents of close friends.

The contribution of this study is three-fold. First, I update the evidence on the importance peers' parental education. Manski (1993b) pioneers the distinction between contemporaneous spillover among peers and the effect of peers' characteristics. The first refers to the 'endogenous effect' caused by peers' outcome and the second refers to the 'contextual' effect generated by peers' pre-existing characteristics.[5] This paper is more related to contextual

[3]Their sample consists of US kindergarteners.

[4]The coefficients are imprecise for students whose mothers are absent, possibly due to a small sample size of this group.

[5]Subsequent work on social interaction pays most of the attention to identify the contemporaneous spillover by modifying the econometric model (Moffitt et al., 2001; Brock and Durlauf, 2001b, 2003; Lee, 2007), random assignment of peers (Sacerdote, 2001; Zimmerman, 2003), or exploiting possible exogenous variations (Hoxby, 2000; Angrist and Lang, 2004; Lavy et al., 2012; Imberman et al., 2012). Most of the work using the AddHealth data also focuses on identifying the contemporaneous spillover (Fletcher, 2010;

effects generated by the family background of peers. Among the studies using the AddHealth data, Bifulco et al. (2011) and Bifulco et al. (2014) analyze cohort composition of college educated mothers and find small long-term effect on educational outcomes. More recently, Cools et al. (2019) define high-achievers using parental education (at least one parent with a degree). They find that the exposure to high-achieving boys has a negative effect on a girl's educational attainment. Studies using other data sources also find positive spillover from peers' parents.[6] Carrell and Hoekstra (2010) find that students suffering from domestic violence create negative spillover to the classroom. Black et al. (2013) analyze random variations of peer composition across cohorts in Norweigan schools and find significant effects of father's earnings of grade-mates on male students. Studies in other countries also find that peers' parents matters (McEwan, 2003; Haraldsvik and Bonesrønning, 2014).[7] For example, Ammermueller and Pischke (2009) exploit random allocations of students into classes and find that the number of books in classmates' homes increases the reading test scores of 9- and 10-year-olds in six European countries. In China, Eble and Hu (2019) find that gender bias of classmates' parents negatively affects the math score of a girl. This paper complements the studies on the effect of peers' parental education in three aspects. First, I examine the friendship network. Methodologically, I estimate the spatial autoregressive model that allows students to have their own peer groups in a network (Bramoullé et al., 2009; Lee et al., 2010; Lin, 2010). This is different from the linear-in-means model where a student is assumed to be affected by all members in a group (school, grade, or class) equally.[8] Second, I differentiate the effects of peers' fathers from that of peers' mothers. Third, I relax the assumption that

Liu et al., 2014; Hsieh and Lee, 2016; Hsieh and Lin, 2017; Patacchini et al., 2017).

[6]There are some exceptions. Boisjoly et al. (2006), who find that the exposure to roommates from a higher-income family does not change the racial attitude on white students. Fruehwirth (2013) finds that better parental education of classmates, reported by teachers, has a negative effect on the reading score of nonwhite students in North Carolina public school.

[7]McEwan (2003) finds a positive spillover from the classroom composition of mother's education on students' performance on Spanish and math in Chile. In Norway, Haraldsvik and Bonesrønning (2014) find that the exposure to classmates from a lower educated family affects boys' achievement.

[8]The non-linearity introduced in the spatial autoregressive model (with network fixed effect) also circumvents the 'reflection problem' to separately identify the 'endogenous effect' and 'contextual effect'. More discussion in Section 3.1.

the effect of peers' parents on educational outcomes is the same for male and female students.

The heterogeneous effect by gender further relates to two growing strands in the economics literature. I provide evidence on when role modeling may take place. Earlier work has acknowledged that non-parental adults affect young people through demonstrating "know-how" of certain behaviors (Borjas, 1995; Wilson, 2012). Bisin and Verdier (2000) shows that the influence of non-parental adults dominates parental influence when socialization within the family is not successful. Later evidence on role modeling typically refers to teachers (Bettinger and Long, 2005; Dee, 2007; Griffith, 2014; Eble and Hu, 2017) or senior peers (Porter and Serra, 2019; Kofoed and mcGoveny, 2019). With the angle of social networks, peers' parents can also demonstrate as role models either through direct interaction or the spread of information (Chung, 2000). My finding indicates that non-parental role models become more relevant when one's family lacks a role model.

The gendered pattern also speaks to the significance of gender norm in driving economic decisions. Akerlof and Kranton (2000, 2002) pioneer the formulation of gender identity in an economic framework. In their setting, utility-maximizing individuals care about self-image. A cost is incurred if they deviate from the average behaviors of the relevant gender group. Thus, the 'ideal behaviors' create a pressure to conform. Subsequent work also finds empirical ground about the importance of gender norms on economic behaviors (Fernández et al., 2004; Fernandez and Fogli, 2009; Alesina et al., 2013; Bertrand et al., 2015; Giménez-Nadal et al., 2019). For example, Haaland et al. (2018) find that gender norm of a community disseminates across generations and affects the gender employment gap of young people. In my finding, with different educational compositions of fathers and mothers in a network, different beliefs on gender norm are passed down through inter-generation transmission.

2 AddHealth Data

The key to study social network empirically is to identify how individuals are connected. The restricted version of the AddHealth dataset ideally suites this paper because it is a unique database on high-school friendship networks in addition to standard demographic details. It is a longitudinal study consists of students in grades 7-12 in the United States from a nationally representative sample of schools starting from the 1994-95 school year.[9] In the first wave, 90,118 students in the Core sample from 132 schools participate in the In-School questionnaire. In this survey, each student is asked to nominate up to five male and five female friends in the school. Therefore, it is not necessary to assume equal influence across all members in a classroom/school. In the subsequent waves, the friendship networks among tracked students are also recorded. However, as Chandrasekhar and Lewis (2011) point out, truncated networks may result in biased estimates. In order to preserve the network structure, the friendship networks are based on the records in the In-School questionnaire because only a subset of the students is sampled in the subsequent surveys. In the main analysis, a friendship link is defined as directed without the need of consensus. A network is defined as a school, which is naturally derived from the survey design.

The main outcome of interest is a dichotomous variable indicating 'college completion' status. Because of the longitudinal nature of AddHealth, I can take a closer look on the long-term effects of social network on human capital accumulation. The survey follows up the socioeconomic circumstances of 10,258 randomly chosen students from the Core sample in 2008 (Wave IV).[10]

Following the literature, networks which are too small (< 11 students) and too big (> 400 students) are not included in the main analysis.[11] Students with no friend nominations are

[9] The survey is still ongoing with subsequent waves in 1996 (Wave II), 2001 and 2002 (Wave III), 2008 (Wave IV), and 2016 to 2018 (Wave V).

[10] There are total 12,105 students drawn from the Core sample. However, some of them do not complete the In-school questionnaire.

[11] The lower bound is set based on the survey design (Hsieh and Lee, 2016), whereas the upper bound is set based on the speed of convergence of the selection-corrected model. Calvó-Armengol et al. (2009) also provide theoretical arguments on excluding networks with extreme size.

dropped. In total, 28% of the observations in the original sample are discarded. Table A1 in Appendix shows the summary statistics of students' characteristics in the data. Indeed, the mean sample characteristics do not change in a meaningful way with the two selection criterion. For example, about 35% of the sampled students completed college and male-female ratio remains 50-50 consistently. The final sample consists of 7,399 students from 116 networks (schools). On average, each student has approximately five friends.

To measure the influence from friends' parents, I use the unique identifier of each student and match the characteristics of friends. In particular, family spillover is measured by the education background of friends' parents – college attainment of friends' fathers and mothers. Table 1 compares the raw relationship between the characteristics of peers' parents and college completion status of students. For both male and female students with college degrees, their friends' parents whom they met during Grade 7 to 12 tend to be better educated. For example, in the first row, only 21% of the peers' fathers went to college for male students without a degree, compared to 39% for male students with a degree. The difference between the two is also statistically significant. This pattern applies also to the college attainment of peers' mothers. This is suggestive to a positive relationship between the educational background of peers' parents and the college outcome of a student.

3 Empirical Specifications

3.1 Sociomatrix and Baseline Model

To formalize the idea of social interaction in the estimation, a sociomatrix for each network (school) is employed. Define n_s be the number of students in network s, and thus the sociomatrix (D_s) is a n_s-by-n_s square matrix in which the rows represent the students and columns represent their potential friends in the network. Each entry of D_s is a dummy indicator $d_{ij,s}$ equals 1 if i nominates j as friend. Under the assumption that friendship links are directed without consensus, D_s is asymmetric. To examine separate effects by

Table 1: Positive correlation between college graduation and friends' family background

	Male			Female		
	With college	No college	Diff.	With college	No college	Diff.
Family Background of friend						
Friends' Father (College)	0.3942	0.2081	-0.1861***	0.3475	0.1738	-0.1738***
Friends' Mother (College)	0.3835	0.2324	-0.1511***	0.3742	0.2051	-0.1691***
Friends from two-parent family	0.8061	0.7117	-0.0944***	0.7723	0.6879	-0.0844***
Characteristics of friend						
GPA	3.0624	2.7377	-0.3248***	3.0368	2.7226	-0.3142***
Female	0.3920	0.3940	0.0021	0.6608	0.6733	0.0125*
Age	15.0014	15.0090	0.0076	15.0415	15.0444	0.0029
Black	0.1292	0.1508	0.0215**	0.1907	0.2043	0.0136
Other races	0.2284	0.2621	0.0336***	0.2096	0.2551	0.0455***
Observations	1296	2373		1977	2917	

Note: Standard errors in parenthesis; ***, ** and * represent 1%, 5% and 10% significance level respectively.

This table compares the raw relationship between the characteristics of peers' parents and college completion status of students. For both male and female students with college degrees, their friends' parents whom they met during Grade 7 to 12 tend to be better educated. For example, in the first row, only 21% of their peers' fathers went to college for males without degree, compared to 39% for males with degree.

gender, D_s is multiplied element-wise by the male and female indicators to generate two separate sociomatrices, D_s^{male} and D_s^{female}.[12] A baseline model to estimate the influence from peers' parents is to regress the outcome variable on the average characteristics of peers' parents. Formally, define W_g^{male} and W_g^{female} as the row-normalized sociomatrices to obtain the average characteristics. For student i in grade g in network s, the probability of college attainment is expressed as:

$$\begin{aligned} Pr(Y_{isg} = 1) = \Phi\{\beta_{male,FATHER} W_s^{male} FATHER_s + \beta_{male,MOTHER} W_s^{male} MOTHER_s \\ \beta_{female,FATHER} W_s^{female} FATHER_s + \beta_{female,MOTHER} W_s^{female} MOTHER_s \\ + W_s X_s \delta + X_{isg}\phi + \alpha_g + \alpha_s + u_{isg}\} \quad (1) \end{aligned}$$

[12] The operation can be interpreted as dividing D_s into two row segments by gender where

$$D_s = \begin{pmatrix} \widetilde{D}_s^{male} \\ \widetilde{D}_s^{female} \end{pmatrix}$$

which is similar to Hsieh and Lin (2017) that the sociomatrix is divided by blocks.

$FATHER_{js}$ and $MOTHER_{js}$ are vectors of indicators equals 1 if j's father/mother graduates from college. Together with the row-normalized sociomatrices, the first four terms measure the proportion of i's peers' fathers/mothers who are college graduates with separate effects on boys and girls. For example, $\beta_{male,FATHER}$ is the effects on boys from having more peers' fathers who are college graduates. X_{isg} is a vector of i's characteristics which are shown in Table A1. Especially, controlling for family and community characteristics help alleviate the concern that families select neighborhoods. α_g and α_s refer to grade and school fixed effects. The average characteristics of peers $W_s X_{js}$ (GPA, race, gender, age, and single parent indicator) are added to isolate β from direct peer effect. Because the number of friends varies, outdegree (outgoing nominations) by race are included as control variables.

The model in Equation 1 with the use of the socio-matrix is called a spatial autroregressive (SAR) model (Lee, 2007). It is different from the traditional linear-in-means model, where an individual is assumed to be connected with everyone in the same network (a complete network). The difference between the two models becomes more apparent on the issue of the 'reflection problem'. Using the terminology by Manski (1993b), the educational background of peers' parents generates the 'contextual effect'. To isolate the direct spillover among peers, a natural strategy is to include the college attainment rate of peers, which is called the 'endogenous effect'. However, this creates the 'reflection problem' in a linear-in-means model. Consider a common specification where peers are defined as the grade-mates. Because everyone in the same grade is assumed to affect each other, the average college attainment of grade-mates is just a linear function of the educational background of grate-mates' parents. In other words, the model regressing an individual outcome on group outcome and group characteristics is under-identified. In contrast, the SAR model incorporates the intransitive nature of actual social links. That is, even in the same network, an individual is only connected with a small set of neighbors who are connected with other non-overlapping neighbors. Therefore, there exist individual-level variations in peers variables. As Lin (2010) noted, the two sources of social effect in a SAR model can be separately identified because the weighted average of peers' outcomes is no longer collinear with the weighted average of peers' characteristics, where the weight depends on $d_{ij,s}$ in D_s. As discussed in the data section, only a subset of students in Wave I is traced through Wave IV. To preserve the network structure inside a school, GPA of peers is a reasonable substitute for the future college status of peers to isolate the spillover of peers' parents from the direct peer effect. The use of a lagged variable to replace the contemporaneous peers outcome also gets around the reciprocal nature of social interactions (Hanushek et al., 2003).

Although the non-linearity in a SAR model circumvents the 'reflection problem', the model is not free from identification issues.[13] First, the association between individual outcomes and peers variables may just be driven by common unobservables such as teacher attributes. This is called the 'correlated effect' by Manski (1993b). Families also sort into schools and neighborhoods. Therefore, I added neighborhood characteristics and school fixed effects (α_s) to mitigate these concerns (Bramoullé et al., 2009). Second, students select into different types of friendship. That is, the choice variable $d_{ij,s}$ in D_s is correlated with both the outcome and u_{isg}. This is a typical omitted variable bias. In the following, I introduce two distinct strategies to address this concern.

3.2 Selection Correction Model

In the baseline model, the inclusion of network (school) fixed effects addresses the selection bias at the school level. In other words, to obtain unbiased estimates for β in Equation 1, we require the assumption that there is no systematic unobserved heterogeneity within each school. However. because the peers variables are constructed using the actual friendship nomination, I also worry about friendship sorting due to individual unobserved heterogeneity. More specifically, the endogenous variable now is the $d_{ij,s}$ entry, which is the choice variable of each potential link in a network. To formally discuss this problem in an econometric

[13] The probit specification in Equation 1 is another source of non-linearity (Brock and Durlauf, 2001a).

framework, assume an additive structure of the residual in Equation 1 such that:

$$u_{isg} = \xi_{isg} + \epsilon_{isg} \tag{2}$$

where ϵ_{isg} is an idiosyncratic error term. $\xi_{ik,s}$ is a vector of unobserved traits $(e_{i1,s}, \ldots.., e_{i\bar{d},s})$ correlates with both the college outcome and $d_{ij,s}$. $\xi_{ik,s}$ can be any traits unobserved to researchers. For example, high-persistent students tend to have better academic performance. They also tend to avoid low-persistent peers. This resembles the omitted variable bias, in which Equation 1 fails to incorporate common factors $(\xi_{ik,s})$ that simultaneously determine the college outcome and link decisions.

Of course, the variety of friendship formation process is not handful (Jackson, 2010). In this subsection, we attempt to address 'homophily', a common social phenomenon (Lazarsfeld et al., 1954; Carrell et al., 2013). 'Homophily' refers to a same type attraction. Two individuals are more likely to connect if they share similar traits. The traits can either be observed such as gender and race, or unobserved such as taste and personality. We can formalize the idea of homophily in a latent space framework (Hoff et al., 2002). The outcome variable in this framework is $d_{ij,s}$, which is the endogenous variable in Equation 1. It equals 1 if i nominates j as his/her friend. The key determinants of the decision of i are the distances of characteristics between himself/herself and agent j. Assuming a logit form of the link decision, the probability of each link equals 0 or 1 is then:

$$Pr(d_{ij,s}|\psi_{ij,s}) = \left(\frac{1}{1+exp(-\psi_{ij,s})}\right)^{d_{ij,s}} \left(\frac{exp(-\psi_{ij,s})}{1+exp(-\psi_{ij,s})}\right)^{1-d_{ij,s}} \tag{3}$$

where

$$\psi_{ij,s} = \gamma_0 + \gamma_1 \mathbf{X}_{i,s} + \gamma_2 \mathbf{X}_{j,s} + \gamma_3 |age_{i,s} - age_{j,s}| + \gamma_4 |gender_{i,s} - gender_{j,s}| + \gamma_5 |grade_{i,s} - grade_{j,s}| + \sum_{k=1}^{\bar{d}} \gamma_{k+5} |e_{ik,s} - e_{jk,s}|$$

The functional form and the choice of variables in Equation 3 closely follow that of Hsieh and Lee (2016).[14] The link decision, $d_{ij,s}$, depends on four components. $\mathbf{X}_i$ and $\mathbf{X}_j$ contain i's and j's age. The dyad-specific variables (absolute distance of gender, age and grade) capture homophily based on observed characteristics. $|e_{ik,s} - e_{jk,s}|$ captures homophily based on an unobserved characteristic, for example, sociable students like to play with sociable students (Massen and Koski, 2014). The subscript k and the summation sign $\sum_{k=1}^{\bar{d}}$ acknowledge the possibility of homophily on multiple unobserved characteristics. The coefficients of the measures on homophily are expected to be negative, i.e. the more dissimilar the two agents are, the less likely they become friends.

The key of this model is the vector of $\bar{d}$-dimensional latent factors, $\xi_{i,s} = (e_{i1,s}, \ldots.., e_{i\bar{d},s})$. The terms provide a linkage between friendship formation and the college outcome equation. To include these latent factors into the outcome equation, Equation 1 is augmented using the approach of Albert and Chib (1993).[15] Assume y^*_{isg} represents a continuous latent variable underlying the college decision. Incorporating the error structure specified in Equation 2, the augmented outcome equation becomes:

$$Y_{isg} = \begin{cases} 1 & \text{if } \; y^*_{isg} > 0 \\ 0 & \text{if } \; y^*_{isg} \leq 0 \end{cases}$$

[14] A variety of selection-corrected models of this kind have been adopted in recent applied social network analysis (Goldsmith-Pinkham and Imbens, 2013; Chan and Lam, 2014; Griffith, 2017).

[15] The augmentation also helps circumvent the incidental parameter problem in a fixed effect non-linear model by assuming a continuous latent variable.

$$y^*_{isg} = \beta_{male,FATHER} W_s^{male} FATHER_s + \beta_{male,MOTHER} W_s^{male} MOTHER_s$$
$$\beta_{female,FATHER} W_s^{female} FATHER_s + \beta_{female,MOTHER} W_s^{female} MOTHER_s$$
$$+ W_s X_s \delta + X_{isg}\phi + \alpha_g + \alpha_s + \sum_{k=1}^{\bar{d}} \rho_k e_{ik,sg} + \epsilon_{isg} \quad (4)$$

and ϵ_{isg} follows standard normal distribution.[16]

The $\bar{d}$-dimensional latent factors enter Equation 4 as control functions. In other words, under the structure in Equation 3 and 4, the researcher can control for an exhaustive list of omitted variables. Same as the Heckman-selection model, the first stage requires excluded variables to identify the model. Therefore, the dyad-specific characteristics (absolute distance of age, gender, and grade) in Equation 3 are used as exclusion restrictions.[17] The assumption is that similarities in observed characteristics, controlling for the characteristics themselves in the outcome equation, only affect the outcome y^*_{isg} through affecting the link formation. The threat to this identifying assumption is that the characteristics are determined after friendships being formed such as common club activities.[18] Therefore, the observed characteristics are all pre-determined.[19]

From Equation 3 and Equation 4, there are four sets of parameters: outcome parameters ($\theta = \{\beta, \delta, \phi, \alpha_g\}$), network fixed effects (α_s), link formation parameters (γ), and the error-correction terms (ρ). For illustration purpose, define $\Theta = \{\theta, \gamma, \Gamma, \rho\}$, and Z be the variables in the first stage. For each school, the joint likelihood function of $Y^*_s = \{y^*_{isg}, ..., y^*_{n_s sg}\}$ and

[16]The error structure in Equation 2, i.e. $u_{isg} = \sum_{k=1}^{\bar{d}} \rho_k e_{ik,sg} + \epsilon_{isg}$, requires the parametric assumptions that $E[u_{isg}|\xi_{i,s}]$ being linear in $\xi_{i,s}$. Similar to the Heckman selection model, u and ξ follow a joint normal distribution (Hsieh and Lee, 2016). However, only $E[u_{isg}|\xi_{i,s}]$ being linear in $\xi_{i,s}$ is a necessary assumption to identify the model (Olsen, 1980).

[17]Without excluded variables in $\psi_{ij,s}$ in Equation 3, the identification of β relies on the non-linearity of $e_{i,s}$, and this can cause imprecise estimates (Brock and Durlauf, 2003).

[18]Chan and Lam (2014) demonstrated the use of pre-determined common hobby as the exclusion restriction.

[19]Notice that although the model in this paper essentially follows Hsieh and Lee (2016), the indicator 'same race' is not included as this correlates strongly with socioeconomic status.

D_s for the estimation is:

$$
\begin{aligned}
L(Y_s^*, D_s | X_s, Z_s, \xi_s; \Theta) &= \int_{\xi_s} P(Y_s^*, D_s | X_s, Z_s, \xi_s; \Theta) f(\xi_s) d\xi_s \\
&= \int_{\xi_s} P(Y_s^* | D_s, X_s, \xi_s; \Theta, \alpha_s) P(D_s | Z_s, \xi_s; \Theta) f(\xi_s) d\xi_s \quad (5)
\end{aligned}
$$

where $P(D_s|Z_s, \xi_s) = \prod_i^{n_s} \prod_{j \neq i}^{n_s} Pr(d_{ij,s}|Z_{i,j,s}, \xi_{i,s})$ under the assumption that each link is formed independently conditional on $Z_{i,j,s}$ and $\xi_{i,s}$. An obvious way to obtain the estimates is to apply Maximum Likelihood to Equation 5. However, with the presence of the unobserved latent factors $\xi_{i,s}$, there is no closed-form solutions. Also, the key component, ξ, cannot be observed by researchers.

To circumvent these difficulties, I estimate Equation 3 and Equation 4 simultaneously using Bayesian method (mixing Metropolis-Hastings (M-H) and Gibbs sampler).[20] Details of MCMC algorithm are described in D. The part that deals with the unobservables, however, is worth further explanations. In each iteration, ξ_i are drawn randomly from a prior distribution and M-H algorithm are used to decide should the draw be updated. This procedure allows researchers to treat the latent factors as if they are known. The chosen draw is then used to update the rest of the parameters.

Goldsmith-Pinkham and Imbens (2013) are among the first to implement this selection-correction approach to look at spillover among peers in recreational activities using a binary latent factor, but finding ρ insignificant in their case. Hsieh and Lee (2016) and Hsieh and Lin (2017) allow the latent factors to be continuous and multidimensional. They show that upon including sufficient dimensions, selection based on homophily can be solved. To decide the optimal dimension of latent factors, I follow their approach to use Akaike's information criterion for Monte Carlo (AICM) introduced by Raftery et al. (2007) for Bayesian model selections. Following Hsieh and Lee (2016), I run 150,000 iterations with the first 40,000 as burn-in.[21]

[20] As noted by Hsieh and Lee (2016), this approach is essentially full information maximum likelihood.
[21] The computational burden of this approach is high. It takes about 70 hours to obtain the full chain.

3.3 Alternative Specification

In Equation 1, the inclusion of network (school) fixed effects and neighborhood characteristics mitigate the concern of sorting across schools and neighborhoods. In Section 3.2, the semi-structural approach further addresses friendship formation caused by homophily within a network (school). However, one may still worry that the selection-correction approach cannot take into account other forms of friendship sorting, for example, all students strategically play with the smartest schoolmates because of taste or parental manipulation.

To check the possibility of other selection mechanisms in driving my finding, I adopt an alternative identification strategy by exploiting the cohort variation in parental characteristics within a school. Formally, I employ a linear-in-means model as follows:

$$\begin{aligned} Y_{i,s,g} = {} & \beta_{male,FATHER}\overline{FATHER}_{-i,g,s} * Male_i + \beta_{male,MOTHER}\overline{MOTHER}_{-i,g,s} * Male_i \\ & \beta_{female,FATHER}\overline{FATHER}_{-i,g,s} * Female_i + \beta_{female,MOTHER}\overline{MOTHER}_{-i,g,s} * Female_i \\ & + W_s X_s \delta + X_{i,s,g}\phi + \alpha_s + \alpha_g + \alpha_s * \tilde{t} + u_{i,s,g} \quad (6) \end{aligned}$$

Compare to the friendship analysis, there are several notable differences in the empirical estimation. First, peers' maternal/paternal education is now measured by the leave-one-out cohort mean of the college status of grade-mates mothers/fathers. Second, the inclusion of school-specific trends, $\alpha_s * \tilde{t}$, alleviates the concern that parents may select schools based on systematic patterns over time, for example, the area has been becoming more urbanized and higher income households are moving in.[22] In terms of the sample, I now can include the students without any friendship nomination. I drop the grades which are too small or too big, corresponding to the 5th or 95th percentile of the grade-size distribution. The final sample consists of 9,716 students from 368 school-by-grade units. The minimum and maximum grade size are 11 and 301 students respectively.

[22] $\tilde{t} = g - 7$, where $g = \{7, 8, 9, 10, 11, 12\}$

This approach is well-established in the peer effect literature (Hoxby, 2000; Bifulco et al., 2011; Olivetti et al., 2018). The identifying assumption is that peer group formation is orthogonal to the difference in the cohort composition of parental education, conditional on the covariates and school fixed effects. This reduced-form approach does not require parametric assumption on a friendship formation model and relies less on the computational power to obtain a stable value chain of the parameters, which are the merits over the selection-correction model. However, the peer group under this alternative approach is defined as the grade-mates. This losses valuable friendship information such as the closeness of friends to explore possible mechanisms. Therefore, I adopt the linear-in-means model with cohort differences in parental composition as a complementary strategy to the selection-correction approach.

4 Results

4.1 The Role of Control Variables

In the baseline specification, I decompose the effect of peers' parental education by gradually adding different sets of covariates. I first present the results of homogeneous effects which are the pooled estimates of the effects on both male and female students.

In Column 1 of Table 2, without any control variable, there exist a strong and positive relationship between the average college attainment of peers' parents and the college attainment of a student. This is the pure correlation observed in the summary statistics in Table 1. An interesting observation is that the size of the effect of peers' paternal education is twice as large as that peers' maternal education even though both are estimated with a similar precision. The difference is also statistically significant at 1% level. The inclusion of own characteristics (GPA, age, race, and gender) does not cause a big change in the magnitude as shown in Column 2.

When own family background and neighborhood characteristics are included, the size of

Table 2: Control variables reduce the size of the effects from peers' parents

	(1)	(2)	(3)	(4)	(5)	(6)
Peers' Paternal Education	1.111***	1.012***	0.735***	0.620***	0.506***	0.297***
	(0.0642)	(0.0694)	(0.0726)	(0.0741)	(0.0808)	(0.0863)
Peers' Maternal Education	0.677***	0.567***	0.392***	0.315***	0.186**	0.118
	(0.0633)	(0.0675)	(0.0702)	(0.0711)	(0.0780)	(0.0793)
F-Stat of Difference	15.19***	13.76***	7.64***	5.96**	5.73**	1.69
Constant	-0.774***	-3.597***	-3.801***	-4.203***	-0.745	-1.230
	(0.0233)	(0.174)	(0.182)	(0.190)	(0.637)	(0.796)
Own Characteristics		X	X	X	X	X
Family Background#			X	X	X	X
Neighborhood Characteristics				X	X	X
Fixed Effects					X	X
Peers' Controls##						X
Pseudo R2	0.0909	0.203	0.253	0.263	0.309	0.318
Observations	7,399	7,399	7,399	7,399	7,352	7,352

Note: Standard errors in parenthesis; ***, ** and * represent 1%, 5% and 10% significance level respectively.

Dependent variable in all regressions is a college completion indicator. A network is defined as a school. Peers' maternal/paternal education is measured by the proportion of peers' mothers/fathers who have a college degree.

Family background include occupation and education of father and mother, and a single parent indicator; neighborhood characteristics include crime rate (county level), median household income (block level) and race dispersion (block level).

Friendship links are directed without consensus. 'Peers' Controls' refer to the average characteristics (GPA, race, gender, age and single parent status) of peers, and outdegree (by race).

the estimates for peers' paternal and maternal education are reduced by 39% and 44% as shown from Column 3 to 4. When grade and the school (network) fixed effects are added in Column 5 to capture network-level unobserved heterogeneity, the magnitudes are further reduced by 18% and 41%.[23] From Column 3 to 5, the changes show the importance of socioeconomic sorting in driving the positive relationship between peers' parental education and a student's achievement. In Column 6, I estimate the fully saturated model that takes direct peer effects into account by including peers' characteristics (including GPA, race, gender, age and single parent status). Both the effects of peers' paternal and maternal education are further reduced by 41% and 37%, and only the former remains statistically significant. Throughout the exercise, the size of the influence of peers' fathers is statistically

[23] A small amount of observations drops out when fixed effects are added due to the incidental parameters problem in non-linear models. This can be circumvented when I estimate the selection-corrected model in Section 3.2 because a continuous latent variable is assumed under the framework of Albert and Chib (1993).

larger than that of peers mothers. However, the difference is explained away by direct peer effects.

The above estimates of the average effect across student's gender are consistent with previous studies which also analyze AddHealth data. Hsieh and Lee (2016) treat the family background of friends as control variables using 'the proportion of peers' mothers who graduate from high-school' and find insignificant effects.[24] In contrast, Patacchini and Zenou (2016) find significant and positive effects from parental education of peers. They define 'parent' as the interviewee in the In-Home survey, which can either be father or mother. Therefore, their estimate is essentially mixing the effects of peers' fathers and peers' mothers.

I repeat the exercise above to examine the role of control variables in explaining the heterogeneous effects by student's gender in Table 3. In Column 1, without control variables, the size of the effects for both same-gender and opposite-gender spillover are large and positive. Same as the homogeneous effect analysis in Table 2, the influence of peers' fathers is larger than that of peers' mothers. The gender-specific pattern on boys has already emerged where the effect of peers' paternal education is statistically larger that the effect of peers' maternal education (f-stat: 21.66). Again, adding own characteristics in Column 2 does not significantly alter the four estimates. In Column 3, we start to see the role of sorting based on socioeconomic status. Once controlling for own family background, the most affected variable is the opposite-gender spillover on boys. The magnitude drops by more than a half and becomes statistically insignificant. In contrast, the size of the same-gender spillover (for both boys and girls) and the effect of peers' paternal education on girls remain statistically significant. This qualitative pattern remains the same when neighborhood characteristics and fixed effects are added in Column 4 and 5. When direct peer effects are included in Column 6, the effect from peers' fathers on female students drops by 59% and becomes insignificant. The same-gender spillovers for male and female students also decrease by 29% and 15%. However, they are the only effects that remain statistically significant.

[24]They use high-school graduation status (HS) as cutoff, and include "less than HS" and "more than HS". The estimate of "more than HS" is essentially close to zero.

Table 3: Family background and direct peer effects explain away the opposite-gender spillover

	(1)	(2)	(3)	(4)	(5)	(6)
Peers' Paternal Education on Boys	1.154***	1.063***	0.826***	0.713***	0.646***	0.460***
	(0.0923)	(0.103)	(0.107)	(0.108)	(0.115)	(0.120)
Peers' Maternal Education on Boys	0.370***	0.383***	0.168	0.0907	-0.0494	-0.142
	(0.0920)	(0.103)	(0.107)	(0.109)	(0.117)	(0.119)
F-Stat of Difference	21.66***	14.17***	12.30***	10.79***	12.13***	8.81***
Peers' Paternal Education on Girls	1.125***	0.976***	0.668***	0.552***	0.395***	0.162
	(0.0857)	(0.0930)	(0.0969)	(0.0984)	(0.106)	(0.111)
Peers' Maternal Education on Girls	0.911***	0.704***	0.557***	0.478***	0.357***	0.305***
	(0.0817)	(0.0893)	(0.0922)	(0.0930)	(0.100)	(0.101)
F-Stat of Difference	2.03	2.92*	0.46	0.20	0.04	0.64
Constant	-0.780***	-3.545***	-3.750***	-4.153***	-0.629	-1.093
	(0.0234)	(0.176)	(0.184)	(0.191)	(0.638)	(0.797)
Own Characteristics		X	X	X	X	X
Family Background#			X	X	X	X
Neighborhood Characteristics				X	X	X
Fixed Effects					X	X
Peers' Controls##						X
Observations	7,399	7,399	7,399	7,399	7,352	7,352
Pseudo R2	0.0956	0.203	0.254	0.264	0.310	0.319

Note: Standard errors in parenthesis; ***, ** and * represent 1%, 5% and 10% significance level respectively.

Dependent variable in all regressions is a college completion indicator. A network is defined as a school. Peers' maternal/paternal education is measured by the proportion of peers' mothers/fathers who have a college degree.

Family background include occupation and education of father and mother, and a single parent indicator; neighborhood characteristics include crime rate (county level), median household income (block level) and race dispersion (block level).

Friendship links are directed without consensus. 'Peers' Controls' refer to the average characteristics (GPA, race, gender, age and single parent status) of peers, and outdegree (by race).

The decomposition exercise here tells the importance of direct peer effects and socioeconomic sorting. In particular, comparing the fully-saturated model in Column 6 and the simple model in Column 2 of Table 3, the two factors explain about half of the same-gender spillover and completely explain away the opposite-gender spillover. However, peers' GPA, as usually used in peer effect studies, does not fully explain the same-gender spillover from peers' parents.

Table 4: Influence Of Peers' Parents Is Smaller Than That Of Peers And Own parents

	Own Parents	Peers' Parents	
	Marginal effect	Marginal effect	Normalized magnitude#
Father on Boys	0.075***	0.118***	0.0236
	(0.0167)	(0.031)	
Mother on Boys	0.085***	-0.037	-0.0074
	(0.0162)	(0.031)	
Father on Girls	0.109***	0.042	0.0084
	(0.0154)	(0.029)	
Mother on Girls	0.104***	0.079***	0.0158
	(0.014)	(0.026)	
Peers' GPA		0.065***	
		(0.011)	

Note: Standard errors in parenthesis; ***, ** and * represent 1%, 5% and 10% significance level respectively.

The magnitudes from peers' parents are adjusted by a factor of 5 as the average number of friends in the sample is 4.85.

The first panel compares the marginal effects from average college attainment of peers' parents to own parents. For a boy, the effect from having one more peers' father with a college degree is one-third of the effect from having a college-grad father. For a girl, the effect from having one more peers' mother with a college degree is one-sixth of the effects from having a college-grad mother.

The second panel shows the marginal effects for average GPA of peers. The influences from peers' parents of same-gender are comparable to half of the effects from an increase in a 1 standard deviation (0.55) increase in average peers' GPA.

Table 4 compares the marginal effects of peers, peers' parents, and own parents on the likelihood of college attainment. To adjust the unit of comparison, the magnitudes from peers' parents are adjusted by a factor of 5 as the average number of friends in the sample is 4.85. For boys, the effect of having one more peers' father with a college degree is one-third of the effect from own parents. For girls, the effect of having one more peers' mother with a college degree is one-sixth of the effect from own parents. The second panel shows the marginal effect of peers' GPA. The same-gender spillover from peers' parents is comparable to half of the effect from an increase in a 1 standard deviation (0.55) increase in average peers' GPA.

4.2 Addressing Friendship Endogeneity

In Table 5, I present the estimates for the selection-corrected model in Section 3.2. The point estimates reported in Table 5 are the mean of the 110,000 posterior draws and hypothesis testing follows frequentist's approach. Convergence is confirmed by Geweke (1992)'s method. The chain values and histograms are presented in D.

Table 5: Gender-specific effects remain robust after selection-correction

	Exogenous link	Endogenous link		
		d-dimensional latent factors		
	(1)	(2)	(3)	(4)
		$d = 1$	$d = 2$	$d = 3$
Peers' Paternal Education on Boys	0.467***	0.470***	0.460***	0.466***
	(0.120)	(0.120)	(0.120)	(0.119)
Peers' Maternal Education on Boys	-0.146	-0.147	-0.162	-0.159
	(0.119)	(0.119)	(0.119)	(0.119)
Peers' Paternal Education on Girls	0.165	0.164	0.151	0.163
	(0.111)	(0.111)	(0.112)	(0.111)
Peers' Maternal Education on Girls	0.307***	0.310***	0.305***	0.299***
	(0.102)	(0.102)	(0.101)	(0.101)
Average GPA of Peers	0.255***	0.254***	0.238***	0.238***
	(0.043)	(0.043)	(0.043)	(0.044)
ρ_1		0.043**	0.063***	0.069***
		(0.018)	(0.017)	(0.017)
ρ_2			0.034**	0.027
			(0.018)	(0.021)
ρ_3				0.023
				(0.019)
Link formation				
$\lvert\text{gender}_i - \text{gender}_j\rvert$		-0.318***	-0.425***	-0.523***
		(0.050)	(0.042)	(0.038)
$\lvert\text{age}_i - \text{age}_j\rvert$		-0.184***	-0.236***	-0.322***
		(0.039)	(0.039)	(0.030)
$\lvert\text{grade}_i - \text{grade}_j\rvert$		-1.311***	-1.361***	-1.341***
		(0.047)	(0.048)	(0.030)
AICM	7015.32	7012.1	7007.716	7008.21

Note: Variables in link formation equation also include i's age, j's age, and the d-dimensional latent factors. MCMC estimation runs for 150,000 iterations with the first 40,000 iterations as burn-in. Standard deviation of the 110,000 posterior draws is in the corresponding parenthesis. Hypothesis testing is based on frequentist's approach, where ***, ** and * represent 1%, 5% and 10% significance level respectively.

Sample size is 7,399. Dependent variable in all regressions is a college completion indicator. All regressions include control variables in Equation 1, grade and school fixed effects. Peers' maternal/paternal education is measured by the proportion of peers' mothers/fathers who have a college degree.

For consistency, I still present the result using Bayesian method for the fully-saturated model to obtain AICM in Column (1) of Table 5. The estimates are essentially the same as that using classical approach in Table 3. From Column (2) to (4), I estimate the selection-corrected model by increasing the dimensions of error correction terms. This exercise is the same as adding more measures of unobserved ability and search the best fit. The first stage results are presented in the second panel which confirm the hypothesis of homophily that two students are less likely to be friends if their pre-determined characteristics are different from each other. According to AICM, the model with two-dimensional latent factors in Column (3) of Table 5 provides the best goodness-of-fit. This result is consistent with the estimation by Hsieh and Lee (2016) in which they also have the best fit of the endogenous friendship model when two-dimensional latent factors are included. The magnitude of the four variables decreases when compared to that in Column (1) and the gender-specific pattern remains robust. However, the drop in magnitudes is not statistically significant.

This does not mean the selection-correction method does not function well. Comparing to the coefficient of the average GPA of peers in Column (1) and (3), the magnitude drops by more than one-third of the standard deviation of the posterior draw. In Figure 1, the CDF of the coefficient on 'Peers' Fathers on Males' (top-left) and 'Peers' GPA' (bottom-right) under the 2-dimension selection-corrected model are more apparent to be first-order stochastically dominated by the CDF under the baseline model.

In the sub-analysis using friendship data, I will employ the selection-corrected model with two-dimensional error-correction terms.

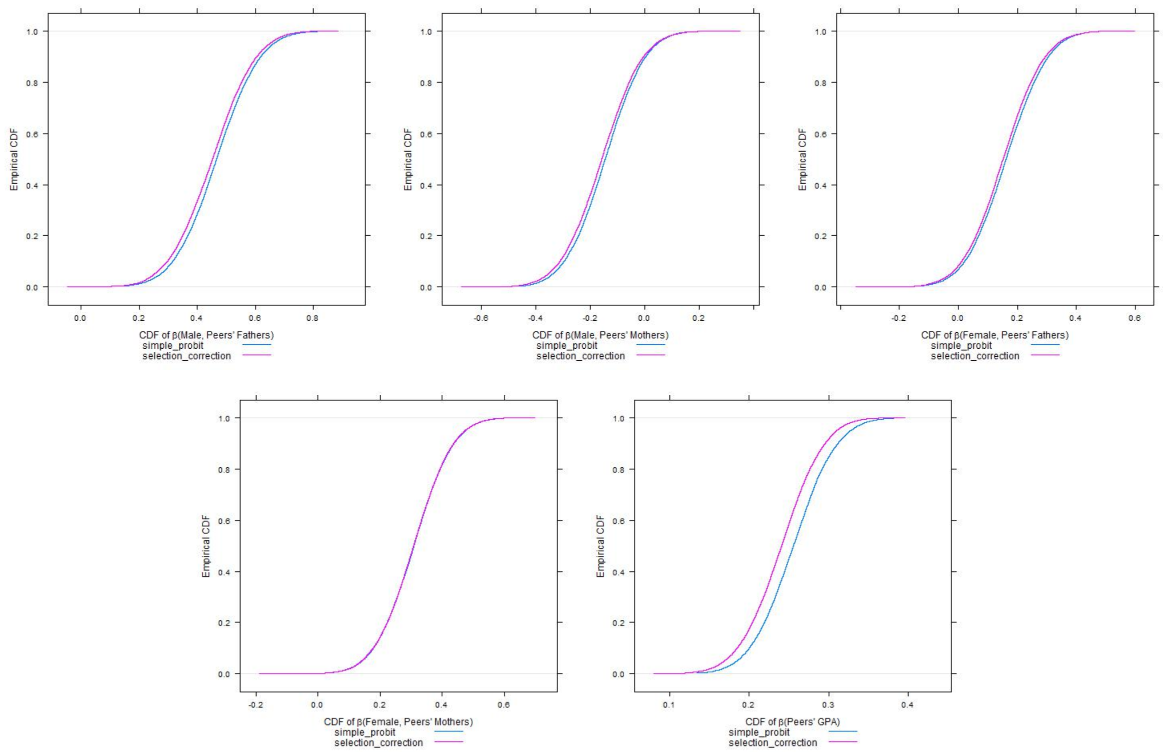

Figure 1: The CDF of the coefficient on 'Peers' Fathers on Males' (top-left) and 'Peers' GPA' (bottom) under the 2-dimension selection-corrected model are more apparent to be first-order stochastically dominated by the CDF under the baseline model.

4.3 Define Peers as Grade-mates

I now present the result for Equation 6, which defines peers as grade-mates. In Column 1 of Table 6, the effect of peers' paternal education is stronger than that of peers' maternal education. Column 2 further shows that the effect of peers' paternal education mainly affects boys. These patterns are qualitatively the same as the finding using the SAR model despite different definitions on peers. However, neither of the four coefficients of interest in the linear-in-means model is statistically significant. From Column 3 to 6, students without friendship nomination are dropped to form a sample comparable to the friendship analysis.[25] Again, I do not observe any significant spillover from peers' parental education, with or without various sets of covariates.

The linear-in-means model seems to give a conflicting result. The only finding that mirrors the friendship analysis is the effect of peers' paternal education on boys. However, in the next section, I will show that both specifications yield the same conclusion on students from a disadvantaged background.

5 Testing Role Model Effect

5.1 Heterogeneity by Family Background

A natural way to test the existence of the role model effect is to categorize friendship links by whether the student interacts with other's parents. The role model effect should only go through the links with the interaction between the student and his/her peer's parents, and the links without the interaction can be used as a placebo test. However, the AddHealth survey lacks this type of information. I therefore consider three alternative tests.

The first alternative is to explore the heterogeneity by family background. The idea is that young people from a disadvantaged background look up more to the role models outside

[25] Because a grade with 10 students or less is dropped in this subsection, there is a small discrepancy in the sample size.

Table 6: Insignificant Spillover Effect When Peers' Refer To Grade-mates

	Including students without a friend		Main Sample			
	(1)	(2)	(3)	(4)	(5)	(6)
Grade-mates' Paternal Education	0.0714		0.0886			
	(0.0621)		(0.0678)			
Grade-mates' Paternal Education	-0.00586		-0.0369			
	(0.0597)		(0.0664)			
Grade-mates' Paternal Education on Boys		0.0857		0.0860	0.0758	0.0465
		(0.0743)		(0.0829)	(0.0826)	(0.107)
Grade-mates' Maternal Education on Boys		-0.0540		-0.0639	-0.0559	-0.162
		(0.0730)		(0.0824)	(0.0819)	(0.0998)
Grade-mates' Paternal Education on Girls		0.0593		0.0911	0.0862	0.0523
		(0.0713)		(0.0777)	(0.0770)	(0.102)
Grade-mates' Maternal Education on Girls		0.0328		-0.0181	-0.0207	-0.112
		(0.0682)		(0.0758)	(0.0749)	(0.0936)
Constant	0.454***	0.472***	0.381**	0.394***	0.332*	0.555*
	(0.132)	(0.132)	(0.148)	(0.149)	(0.175)	(0.305)
Observations	9,716	9,716	7,060	7,060	7,060	7,060
R-squared	0.311	0.311	0.337	0.337	0.347	0.361
Peers' Control					X	X
School Specific Trends						X

Note: Robust standard errors in parentheses; ***, ** and * represent 1%, 5% and 10% significance level respectively.

Dependent variable is a college completion indicator. All regressions include control variables in Equation 1, grade and school fixed effects. Peers' maternal/paternal education refers to the leave-one-out cohort mean of the college status of mothers and fathers respectively. Peers' control refers to the average peers' characteristics (WX) in the main analysis which are constructed based on the actual friendship nomination.

their family (Gustafson et al., 1992; Bisin and Verdier, 2001). The richness of the survey enables me to separate the students into six different categories: families with the presence of both parents are grouped according to whether both, either, or none of the parents are college graduates; together with single father households, single mother households, and households with the absence of both parents.[26] For clarity, a family has a high socioeconomic status (high SES) if both father and mother have a college degree.

In Table 7, the SAR model is modified by interacting the four variables of interest with indicators of the six family status in one regression to capture the heterogeneity. For the effect of peers' paternal education on boys, the effect concentrates on lower SES students. The effect on boys from a single-mother family is slightly smaller and is marginally significant. For the effect of peers' maternal education on girls, the effect is significant only if one of their parents completed college. Two observations suggest the existence of a role model

[26]To determine whether the parents are absent, I make use of the two questions: "Do you live with your biological mother, stepmother, foster mother, or adoptive mother?" and "Do you live with your biological father, stepfather, foster father, or adoptive father?". The number of observations of each group is: Both college (1,237); Either college (1,291); Neither college (2,925); Single FATHER (224); Single MOTHER (1,428); Both absence (294).

Table 7: Lower SES Students Look Up To Parents Of Higher SES Friends - Friendship data

	Two-Parent family					
	Both college	Either college	Neither college	Single FATHER	Single MOTHER	Both absence
Peers' Paternal Education on Boys	0.281	0.580**	0.586***	0.0485	0.507*	0.334
	(0.250)	(0.229)	(0.207)	(0.516)	(0.263)	(0.544)
Peers' Maternal Education on Boys	-0.237	-0.326	0.0533	0.165	0.0328	-0.270
	(0.249)	(0.226)	(0.212)	(0.516)	(0.263)	(0.558)
Peers' Paternal Education on Girls	0.0464	-0.176	0.0846	0.218	0.745***	-0.786
	(0.254)	(0.227)	(0.167)	(0.582)	(0.246)	(0.741)
Peers' Maternal Education on Girls	0.411	0.661***	0.198	0.945	0.0436	-0.0195
	(0.251)	(0.216)	(0.158)	(0.644)	(0.214)	(0.585)

Note: Dependent variable is a college completion indicator. A selection-corrected approach with two-dimensional error-correction terms are estimated. The MCMC estimation runs for 150,000 iterations with the first 40,000 iterations as burn-in. Standard deviation of the 110,000 posterior draws is in the corresponding parenthesis. Hypothesis testing is based on frequentist's approach, where ***, ** and * represent 1%, 5% and 10% significance level respectively.

This table shows the four variables of interest by student's family background. All the variables are estimated by interacting the four interest variables with indicators of family status in one regression. The number of observations of each group is: Both college (1,237); Either college (1,291); Neither college (2,925); Single FATHER (224); Single MOTHER (1,428); Both absence (294). The regression includes control variables in Equation 1, grade and school fixed effects. Peers' maternal/paternal education is measured by the proportion of peers' mothers/fathers who have a college degree.

effect. First, peers' fathers are influential on students from single-mother family. Especially for girls, this is the only family category that I find significant opposite-gender spillover. The size of the effect is also significantly different from that on students from a two-parent family. Second, students from a high SES family do not experience significant effects from peers' parents, regardless of student's gender. The size of the effect is also smaller than those found in lower-educated households, although the difference in magnitudes is not statistically significant.

In Table 8, the linear-in-means model is able to mirror the gendered pattern in the SAR model when the peers parental education is interacting with own family background. I first define an indicator 'Disadvantaged' equals 1 for a low SES family. This definition is based on the finding in Table 7. In Column 1, the gender-specific pattern emerges again only for students from a lower-educated family. The gender-specific effects for disadvantaged students remain significant across different samples and specifications, with a more salient magnitude for boys. In Column 4, although the effect of peers' maternal education on girls is only marginally significant, the effect is stronger for girls than for boys at 10% level.

Table 8: Lower SES Students Look Up To Parents Of Higher SES Friends - Grade-mates

	Including students without a friend	Main Sample		
	(1)	(2)	(3)	(4)
Disadvantaged*(Grade-mates' Paternal Education on Boys)	0.339**	0.439***	0.444***	0.444***
	(0.145)	(0.159)	(0.160)	(0.163)
Disadvantaged*(Grade-mates' Maternal Education on Boys)	-0.00198	-0.0725	-0.107	-0.114
	(0.139)	(0.154)	(0.153)	(0.157)
Disadvantaged*(Grade-mates' Paternal Education on Girls)	-0.000615	0.000831	0.0136	0.0185
	(0.113)	(0.119)	(0.117)	(0.120)
Disadvantaged*(Grade-mates' Maternal Education on Girls)	0.231**	0.245**	0.215*	0.206*
	(0.110)	(0.116)	(0.116)	(0.119)
Disadvantaged	-0.255***	-0.260***	-0.243***	-0.237***
	(0.0279)	(0.0319)	(0.0318)	(0.0323)
Grade-mates' Paternal Education on Boys	-0.178	-0.257*	-0.273*	-0.297*
	(0.139)	(0.151)	(0.151)	(0.168)
Grade-mates' Maternal Education on Boys	-0.0419	0.00603	0.0498	-0.0392
	(0.133)	(0.145)	(0.145)	(0.159)
Grade-mates' Paternal Education on Girls	0.0516	0.0792	0.0683	0.0343
	(0.109)	(0.114)	(0.113)	(0.134)
Grade-mates' Maternal Education on Girls	-0.159	-0.216*	-0.189*	-0.259**
	(0.107)	(0.112)	(0.111)	(0.125)
Constant	0.556***	0.482***	0.357**	0.542*
	(0.136)	(0.152)	(0.180)	(0.320)
Observations	9,716	7,060	7,060	7,060
R-squared	0.312	0.339	0.352	0.366
Peers' Control			X	X
School Specific Trends				X

Note: Robust standard errors in parentheses; ***, ** and * represent 1%, 5% and 10% significance level respectively.
Dependent variable is a college completion indicator. All regressions include control variables in Equation 1, grade and school fixed effects. Peers' maternal/paternal education refers to the leave-one-out cohort mean of the college status of mothers and fathers respectively. Peers' control refers to the average peers' characteristics (WX) in the main analysis which are constructed based on the actual friendship nomination. 'Disadvantaged' equals 1 for a low SES family.

5.2 Relationship Quality With Own Parents

The second alternative to probe the role model effect is to look at whether the magnitude of the spillover from peers' parents diminishes with the intimacy with own parents. The AddHealth data provides unique information for this inquiry. In the In-Home survey of Wave I, students are asked about the relationship with their fathers and mothers. From 1 (not at all) to 5 (very much), students give a rating on how close do they feel to their mother/father and how much do they think she/he cares about them. Altogether there are four responses – two on mother and two on father. I then construct three care indexes using factor analysis: 'Care from both' using all the four responses, 'Care from mother', and 'Care from father'. The analysis is constrained to two-parent families because the questions are skipped for students from single-parent families. The first four rows of Table 9 show the summary statistics for the four survey questions. Students usually have a higher rating on mothers than on fathers.

Table 9: Summary Statistics of the Measures on Parental Care (Two-parent family)

	mean	sd	min	max
Survey Question				
Feel close to mother?	4.470933	.903511	1	5
Mother cares about you?	4.796809	.7074782	1	5
Feel close to father?	3.927013	1.356822	1	5
Father cares about you?	4.362003	1.305317	1	5
Care Index				
Index on both father and mother	.327315	.7712555	-2.44633	.9060649
Index on mother	.082181	.7195385	-3.469914	.4276676
Index on father	.3226724	.7387105	-1.484134	.81809

Note: The responses are recorded in the Wave 1 In-Home survey of AddHealth. The care indexes are obtained using the principal-factor method.

The second panel shows the summary statistics for the three care indexes.

To implement the analysis, I interact the care indexes with the four variables of interest. The coefficients are reported in Table 10. In Column 1, except the interaction on the effect of peers' mothers on males, all the other three interaction terms are negative as expected. That is, the more care a child receives from own parents, the less spillover he or she experiences from peers' parents. The one that is statistically significant is the interaction term on the same gender spillover of males. I further analyze the care indexes specifically on mother and father in Column 2 and 3 respectively. In Column 2, as reflected by the interaction terms, the care from mother does reduce the same gender spillover on both males and females. However, the effect is not significant. What drives the reduction in the spillover in Column 1 is the care from the father as shown in Column 3.

In B, I estimate a simple linear probability model for the ease of interpretation. In Column 1, changing from having the least caring to the most caring parents completely offsets the spillover on males (the index increases from -2.44 to 0.91 as shown by the min and max in Table 9). In Column 3, changing from having the least caring to the most caring father (the index increases from -1.48 to 0.82) completely offset the positive same gender spillover on males.

Table 10: Spillover from Peers' Parents Diminishes With Care From Own Father

	Two-Parent Family		
	(1)	(2)	(3)
VARIABLES	Care from both	Care from mom	Care from dad
CareIndex*(Peers' Paternal Education on Boys)	-0.495**	-0.232	-0.496**
	(0.194)	(0.209)	(0.202)
CareIndex*(Peers' Maternal Education on Boys)	0.152	0.265	0.0796
	(0.188)	(0.201)	(0.192)
CareIndex*(Peers' Paternal Education on Girls)	-0.110	0.0562	-0.121
	(0.154)	(0.177)	(0.159)
CareIndex*(Peers' Maternal Education on Girls)	-0.104	-0.0110	-0.124
	(0.143)	(0.175)	(0.147)
CareIndex	0.271***	0.0549	0.253***
	(0.0593)	(0.0463)	(0.0601)
Peers' Paternal Education on Boys	0.736***	0.536***	0.729***
	(0.170)	(0.146)	(0.170)
Peers' Maternal Education on Boys	-0.262	-0.233	-0.230
	(0.163)	(0.144)	(0.162)
Peers' Paternal Education on Girls	0.0276	0.000961	0.0336
	(0.136)	(0.129)	(0.137)
Peers' Maternal Education on Girls	0.423***	0.389***	0.430***
	(0.124)	(0.120)	(0.124)

Note: Dependent variable is a college completion indicator. A selection-corrected approach with two-dimensional error-correction terms are estimated. The MCMC estimation runs for 150,000 iterations with the first 40,000 iterations as burn-in. Standard deviation of the 110,000 posterior draws is in the corresponding parenthesis. Hypothesis testing is based on frequentist's approach, where ***, ** and * represent 1%, 5% and 10% significance level respectively. All regressions include control variables in Equation 1, grade and school fixed effects.

Four measures about parental cares are obtained from "How close do you feel to your mother/father?" and "How much do you think she/he cares about you?". The responses are recorded in the Wave 1 In-Home survey of AddHealth. 'CareIndex' is obtained by analyzing the correlation matrix of the measures using the principal-factor method. The 'CareIndex' in Column 1 is obtained using all the four measures, whereas that in Column 2 and 3 are obtained using mother(father)-specific measures. The sample is constrained to two-parent families (5,453 students) because responses are skipped for students from single parent families.

5.3 Closer Peers Matter

The third way to test the role model effect is to consider the relative magnitude by the rank of friends, i.e. the closeness of friends serves as a proxy for the possible interaction with peer's parents. The rank of friends is a unique but under-utilized aspect of the AddHealth data. In the survey, students are explicitly asked to "list your best male friend first, then your next best friend, and so on.". This is the same for female friends. In the empirical framework, given the median number of friends is five, I group the first three best male and female friends as 'close friends' and the rest as 'less close friends'. Table 11 reports the eight coefficients, which are estimated simultaneously in a single regression. Among the 'close' friends in Column 1, boys are disproportionately affected by peers' fathers when comparing the probit coefficient between the effect of peers' fathers (0.365) and peers' mothers (-0.140) on boys. Peers' mothers also disproportionately affect girls when comparing the coefficient between the effect of peers' mothers on boys (-0.140) and girls (0.233). In contrast, in Column 2, although the estimates appear to resemble a gender-specific pattern, the same-gender spillovers on boys and girls are insignificant and smaller than those generated by 'close' friends.

In C, I further decompose the effect by each of the friends. The gender-specific effect on boys are mainly generated by the first two best male friends, whereas the gender-specific effect on girls mainly comes from the third best male friend and the first best female friend. One counter-intuitive pattern is the nonlinear effect from male friends on girls, i.e. maternal education of the first best male friend has zero effect on girls. A possible explanation is that romantic relationship has an adverse effect on girls' educational attainment (Sabia and Rees, 2012).

Table 11: The Gender Specific Pattern Concentrates On Close Friends

	Close Friends	Less Close Friends
Peers' Paternal Education on Boys	0.365***	0.126
	(0.109)	(0.103)
Peers' Maternal Education on Boys	-0.140	0.0362
	(0.109)	(0.104)
Peers' Paternal Education on Girls	0.152	0.0363
	(0.102)	(0.0912)
Peers' Maternal Education on Girls	0.233**	0.140
	(0.0936)	(0.0853)

Note: Dependent variable is a college completion indicator. A selection-corrected approach with two-dimensional error-correction terms are estimated. The MCMC estimation runs for 150,000 iterations with the first 40,000 iterations as burn-in. Standard deviation of the 110,000 posterior draws is in the corresponding parenthesis. Hypothesis testing is based on frequentist's approach, where ***, ** and * represent 1%, 5% and 10% significance level respectively. All coefficients are estimated simultaneously in one regression. The regression includes control variables in Equation 1, grade and school fixed effects.

'Close friends' are defined by the first three best male friends and the first three best female friends.

Putting the pieces above together, the gender-specific spillover from peers' parents is stronger for disadvantaged students and among closer connections. These further reaffirm the existence of a role model effect.

6 Sensitivity Check

In this section, I check the sensitivity of the above results by varying the definition of friendships. In the main analysis using friendship data, a link is directed without consensus. That is, the sample pools both reciprocal and non-reciprocal links. Reciprocal links may involve more opportunities for influence and are more relevant to the role model hypothesis. However, disregarding entirely the non-reciprocal links is erroneous. As pointed out by Clifton et al. (2009) and Geven et al. (2013), non-reciprocal links defined by outward nominations imply the direction of attention. Therefore, dropping the non-reciprocal links loses important information. In fact, when non-reciprocal links are taken away, there are 6,738 observations remain with only 40% of the links preserved. Also, the presence of non-reciprocal links is possible to be caused by measurement errors (Patacchini and Venanzoni, 2014; Patacchini et al., 2017). There are various sources of measurement errors including survey design and respondent's interpretation of friendship (Wang et al., 2012).

Table 12: Sensitivity Of Different Definitions On Friendship

	(1) Half Weight on Non-Reciprocal	(2) Only Reciprocal
Peers' Paternal Education on Boys	0.410***	0.081
	(0.118)	(0.097)
Peers' Maternal Education on Boys	-0.157	0.058
	(0.116)	(0.094)
Peers' Paternal Education on Girls	0.101	0.113
	(0.109)	(0.088)
Peers' Maternal Education on Girls	0.315***	0.232***
	(0.099)	(0.083)

Note: Dependent variable is a college completion indicator. A selection-corrected approach with two-dimensional error-correction terms are estimated. The MCMC estimation runs for 150,000 iterations with the first 40,000 iterations as burn-in. Standard deviation of the 110,000 posterior draws is in the corresponding parenthesis. Hypothesis testing is based on frequentist's approach, where ***, ** and * represent 1%, 5% and 10% significance level respectively. All regressions include control variables in Euqation 1, grade and school fixed effects.

In Column 1 of Table 12, I first consider an intermediate solution by assigning a weight of 0.5 to the non-reciprocal links.[27] This strategy acknowledges the argument that non-reciprocal friendship has smaller but non-negligible influence (Lin and Weinberg, 2014). Both the magnitudes of the same-gender spillover on boys and girls change slightly and

[27]In the first stage of the selection model, d_{ij} still equals 1 for a non-reciprocal link.

the effects remain robust. When only reciprocal links are counted in Column 2, the same-gender spillover on boys becomes small and insignificant. One explanation is that we do lose important information when dropping the non-reciprocal links. Nonetheless, the effect of peers' maternal education on girls remains salient. Although the magnitude is smaller than the estimate in Column 1, the difference is not statistically significant.

7 Interpreting the Gender-Specific Spillover

Previous studies on role modeling typically assume direct interactions with the role models, for example teachers and mentors. An ideal setting where the current study can test the role model effect is to exploit the heterogeneity by the time spent with (or knowledge about) peers' parents in which my data lacks this type of information. The heterogeneity by the closeness of friends is a reasonable proxy for the frequency of contacts. The heterogeneity by family background also suggests that role models who are non-family members are particular influential if own family lacks one. These pieces of evidence support the existence of a role model effect.

Even without face-to-face interaction, a role model effect of peers' parents may operate in a subtle way through the spread of information. This type of role model effect is generated by the 'informational role models', who provide a realized value of current decisions (Manski, 1993a; Chung, 2000). Young people learn about the outcomes of older people and make the effort accordingly. The information from the exceptional ones of their social group (gender) is particularly relevant to their current decisions. The analysis of gender identity by Akerlof and Kranton (2000, 2002) also shares a similar theoretical argument. When social influence is present, an individual conforms to the 'appropriate' standards to avoid losing utility. The gender 'identity' creates an extra cost in utility maximization in which the cost is generated via the deviation from the 'ideal' behaviors.

Relating the two theories to my finding. Peers' parents generate a role model effect in a way that the educational composition of mothers and fathers in a network changes the standard for girls and boys respectively. Different beliefs and attitudes on education prevail among social ties and affect the educational decision of the students. In fact, previous studies have shown that the exposure to inspirational content is sufficient to generate a role model effect (Chong and Ferrara, 2009; Riley et al., 2017). For example, Beaman et al. (2012) find that a greater female representation in local councils changes girls' aspiration and increases their educational attainment in India. A recent study by Chetty et al. (2018) also shows a strong association of black boy's future income with the presence of black fathers at the census-tract level. Therefore, the role model phenomenon stands to reason for peers' parents even though real-life interaction may not be a usual occasion.

8 Concluding Remark

The quality of friendship detail documented in the AddHealth dataset offers a unique opportunity to narrow down the mechanisms that drive the social influence. However, using actual links imposes a challenge to obtain a causal interpretation. In this paper, I attempt to control for friendship selection based on similarities, which is a common phenomenon by human nature. I also attempt an alternative strategy using within-school cohort variations in parental composition. The results from both methods are qualitatively the same for students from a lower socioeconomic status: peers' paternal education affects boys and peers' maternal education affects girl. This gendered pattern is consistent with the role modeling channel, which may involve direct interactions or through the spread of gender-specific information.

I have identified peers' parents as separate agents who also generate social effects. The existence of their external effects, in which many peer effect studies have overlooked, has meaningful implications to two policies regarding child development. The first type of policy involves changing neighborhood compositions. One important argument of moving disadvantaged children to a better neighborhood is the exposure to surrounding environments (Chetty et al., 2016b; Chetty and Hendren, 2018; Chyn, 2018). For example, a large government poverty program in the US called the "Moving-to-Opportunity", which randomly allocates poor households to lower-poverty areas, is found to positively affect children outcomes (Katz et al., 2001; Chetty et al., 2016a). Social ties and the potential role modeling effect by the surrounding adults can be an important mechanism in supporting this type of neighborhood policy.

My results also speak to the policy on classroom tracking (Garlick, 2018). To the extent that children form most of their friendships with their classmates, grouping students by ability can have the unintended consequence of reducing the spillover from peers' parents on disadvantaged students. Therefore, there is a trade-off between the peer effects induced by grade tracking and the indirect effects from peers' parents. My work quantifies the latter to allow for sharper tracking policy that fully accounts for this trade-off. More importantly, the null effect I find for students from better-educated families suggests that mixing individuals from diverse backgrounds may not be a zero-sum game.

References

Akerlof, George A and Rachel E Kranton, "Economics and identity," *Quarterly Journal of Economics*, 2000, *115* (3), 715–753.

_ **and** _ , "Identity and schooling: Some lessons for the economics of education," *Journal of Economic Literature*, 2002, *40* (4), 1167–1201.

Albert, James H and Siddhartha Chib, "Bayesian analysis of binary and polychotomous response data," *Journal of the American Statistical Association*, 1993, *88* (422), 669–679.

Alesina, Alberto, Paola Giuliano, and Nathan Nunn, "On the origins of gender roles: Women and the plough," *The Quarterly Journal of Economics*, 2013, *128* (2), 469–530.

Ammermueller, Andreas and Jörn-Steffen Pischke, "Peer effects in European primary schools: Evidence from the progress in international reading literacy study," *Journal of Labor Economics*, 2009, *27* (3), 315–348.

Angrist, Joshua D and Kevin Lang, "Does school integration generate peer effects? Evidence from Boston's Metco Program," *American Economic Review*, 2004, *94* (5), 1613–1634.

Beaman, Lori, Esther Duflo, Rohini Pande, and Petia Topalova, "Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India," *Science*, 2012, *335* (6068), 582–586.

Bertrand, Marianne, Emir Kamenica, and Jessica Pan, "Gender identity and relative income within households," *The Quarterly Journal of Economics*, 2015, *130* (2), 571–614.

Bettinger, Eric P and Bridget Terry Long, "Do faculty serve as role models? The impact of instructor gender on female students," *American Economic Review*, 2005, *95* (2), 152–157.

Bifulco, Robert, Jason M Fletcher, and Stephen L Ross, "The effect of classmate characteristics on post-secondary outcomes: Evidence from the Add Health," *American Economic Journal: Economic Policy*, 2011, *3* (1), 25–53.

_ **,** _ **, Sun Jung Oh, and Stephen L Ross**, "Do high school peers have persistent effects on college attainment and other life outcomes?," *Labour economics*, 2014, *29*, 83–90.

Bisin, Alberto and Thierry Verdier, "Beyond the melting pot: cultural transmission, marriage, and the evolution of ethnic and religious traits," *Quarterly Journal of Economics*, 2000, *115* (3), 955–988.

_ **and** _ , "The economics of cultural transmission and the dynamics of preferences," *Journal of Economic Theory*, 2001, *97* (2), 298–319.

Black, Sandra E, Paul J Devereux, and Kjell G Salvanes, "Under pressure? The effect of peers on outcomes of young adults," *Journal of Labor Economics*, 2013, *31* (1), 119–153.

Boisjoly, Johanne, Greg J Duncan, Michael Kremer, Dan M Levy, and Jacque Eccles, "Empathy or antipathy? The impact of diversity," *American Economic Review*, 2006, *96* (5), 1890–1905.

Borjas, George J, "Ethnicity, Neighborhoods, and Human-Capital Externalities," *American Economic Review*, 1995, *85* (3), 365.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin, "Identification of peer effects through social networks," *Journal of Econometrics*, 2009, *150* (1), 41–55.

Brock, William A and Steven N Durlauf, "Discrete choice with social interactions," *The Review of Economic Studies*, 2001, *68* (2), 235–260.

_ and _ , "Interactions-based models," *Handbook of Econometrics*, 2001, *5*, 3297–3380.

Brock, William and Steven N Durlauf, "Multinomial Choice with Social Interactions," 2003.

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

Carrell, Scott E and Mark L Hoekstra, "Externalities in the classroom: How children exposed to domestic violence affect everyone's kids," *American Economic Journal: Applied Economics*, 2010, *2* (1), 211–28.

_ , Bruce I Sacerdote, and James E West, "From natural variation to optimal policy? The importance of endogenous peer group formation," *Econometrica*, 2013, *81* (3), 855–882.

Chan, Tszkin Julian and Chungsang Tom Lam, "Type of Peers Matters: A Study of Peer Effects of Friends, Studymates and Seatmates on Academic Performance," 2014.

Chandrasekhar, Arun and Randall Lewis, "Econometrics of sampled networks," 2011.

Chetty, Raj and Nathaniel Hendren, "The impacts of neighborhoods on intergenerational mobility I: Childhood exposure effects," *Quarterly Journal of Economics*, 2018, *133* (3), 1107–1162.

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

_ , _ , Frina Lin, Jeremy Majerovitz, and Benjamin Scuderi, "Childhood environment and gender gaps in adulthood," *American Economic Review*, 2016, *106* (5), 282–88.

_ , _ , Maggie R Jones, and Sonya R Porter, "Race and economic opportunity in the United States: An intergenerational perspective," 2018.

Chong, Alberto and Eliana La Ferrara, "Television and divorce: Evidence from Brazilian novelas," *Journal of the European Economic Association*, 2009, *7* (2-3), 458–468.

Chung, Kim-Sau, "Role models and arguments for affirmative action," *American Economic Review*, 2000, *90* (3), 640–648.

Chyn, Eric, "Moved to opportunity: The long-run effects of public housing demolition on children," *American Economic Review*, 2018, *108* (10), 3028–56.

Clifton, Allan, Eric Turkheimer, and Thomas F Oltmanns, "Personality disorder in social networks: Network position as a marker of interpersonal dysfunction," *Social networks*, 2009, *31* (1), 26–32.

Cools, Angela, Raquel Fernández, and Eleonora Patacchini, "Girls, Boys, and High Achievers," Technical Report, National Bureau of Economic Research 2019.

Dee, Thomas S, "Teachers and the gender gaps in student achievement," *Journal of Human Resources*, 2007, *42* (3), 528–554.

Eble, Alex and Feng Hu, "Stereotypes, Role Models, and the Formation of Beliefs," 2017.

_ **and** _ , "How important are beliefs about gender differences in math ability? Transmission across generations and impacts on child outcomes," *CDEP-CGEG Working Paper*, 2019, *53.*

Fernández, Raquel, Alessandra Fogli, and Claudia Olivetti, "Mothers and sons: Preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 2004, *119* (4), 1249–1299.

Fernandez, Raquel and Alessandra Fogli, "Culture: An empirical investigation of beliefs, work, and fertility," *American economic journal: Macroeconomics*, 2009, *1* (1), 146–77.

Fletcher, Jason M, "Social interactions and smoking: Evidence using multiple student cohorts, instrumental variables, and school fixed effects," *Health Economics*, 2010, *19* (4), 466–484.

Fruehwirth, Jane Cooley, "Identifying peer achievement spillovers: Implications for desegregation and the achievement gap," *Quantitative Economics*, 2013, *4* (1), 85–124.

_ **and Jessica Gagete-Miranda**, "Your peers parents: Spillovers from parental education," *Economics of Education Review*, 2019, *73*, 101910.

Garlick, Robert, "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment.," *American Economic Journal: Applied Economics*, 2018, *10* (3), 345–69.

Geven, Sara, Jeroen Weesie, and Frank Van Tubergen, "The influence of friends on adolescents behavior problems at school: The role of ego, alter and dyadic characteristics," *Social Networks*, 2013, *35* (4), 583–592.

Geweke, John, "Evaluating the accuracy of sampling-based approaches to the calculations of posterior moments," *Bayesian statistics*, 1992, *4*, 641–649.

Giménez-Nadal, J Ignacio, Lucia Mangiavacchi, and Luca Piccoli, "Keeping inequality at home: The genesis of gender roles in housework," *Labour Economics*, 2019, *58*, 52–68.

Goldsmith-Pinkham, Paul and Guido W Imbens, "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, 2013, *31* (3), 253–264.

Granovetter, Mark S, "The Strength of Weak Ties," *American Journal of Sociology*, 1973, *78* (6), 1360–1380.

Griffith, Alan, "Random assignment with non-random peers: A structural approach to counterfactual treatment assessment," 2017.

Griffith, Amanda L, "Faculty gender in the college classroom: Does it matter for achievement and major choice?," *Southern Economic Journal*, 2014, *81* (1), 211–231.

Gustafson, Sigrid B, Hakan Stattin, and David Magnusson, "Aspects of the development of a career versus homemaking orientation among females: The longitudinal influence of educational motivation and peers," *Journal of Research on Adolescence*, 1992, *2* (3), 241–259.

Haaland, Venke Furre, Mari Rege, Kjetil Telle, and Mark Votruba, "The intergenerational transfer of the employment gender gap," *Labour Economics*, 2018, *52*, 132–146.

Hanushek, Eric A, John F Kain, Jacob M Markman, and Steven G Rivkin, "Does Peer Ability Affect Student Achievement?," *Journal of Applied Econometrics*, 2003, pp. 527–544.

Haraldsvik, Marianne and Hans Bonesrønning, "Peer Effects on Student Achievement: Does The Education Level of Your Classmates Parents Matter," Technical Report, Working Paper 2014.

Heckman, James J, "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," in "Annals of Economic and Social Measurement, Volume 5, number 4," NBER, 1976, pp. 475–492.

_ , "Sample Selection Bias as a Specification Error," *Econometrica*, 1979, *47* (1), 153–161.

Hoff, Peter D, Adrian E Raftery, and Mark S Handcock, "Latent space approaches to social network analysis," *Journal of the American Statistical Association*, 2002, *97* (460), 1090–1098.

Hoxby, Caroline, "Peer effects in the classroom: Learning from gender and race variation," Technical Report, National Bureau of Economic Research 2000.

Hsieh, Chih-Sheng and Lung-Fei Lee, "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, 2016, *31* (2), 301–319.

— **and Xu Lin**, "Gender and racial peer effects with endogenous network formation," *Regional Science and Urban Economics*, 2017, *67*, 135–147.

Imberman, Scott A, Adriana D Kugler, and Bruce I Sacerdote, "Katrina's children: Evidence on the structure of peer effects from hurricane evacuees," *American Economic Review*, 2012, *102* (5), 2048–82.

Jackson, Matthew O, *Social and Economic Networks*, Princeton university press, 2010.

— **and Asher Wolinsky**, "A strategic model of social and economic networks," *Journal of economic theory*, 1996, *71* (1), 44–74.

Katz, Lawrence F, Jeffrey R Kling, and Jeffrey B Liebman, "Moving to opportunity in Boston: Early results of a randomized mobility experiment," *Quarterly Journal of Economics*, 2001, *116* (2), 607–654.

Kofoed, Michael S and Elizabeth mcGoveny, "The Effect of Same-Gender or Same-Race Role Models on Occupation Choice Evidence from Randomly Assigned Mentors at West Point," *Journal of Human Resources*, 2019, *54* (2), 430–467.

Lavy, Victor, Olmo Silva, and Felix Weinhardt, "The good, the bad, and the average: Evidence on ability peer effects in schools," *Journal of Labor Economics*, 2012, *30* (2), 367–414.

Lazarsfeld, Paul F, Robert K Merton et al., "Friendship as a social process: A substantive and methodological analysis," *Freedom and Control in Modern Society*, 1954, *18* (1), 18–66.

Lee, Lung-Fei, "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 2007, *140* (2), 333–374.

— **, Xiaodong Liu, and Xu Lin**, "Specification and estimation of social interaction models with network structures," *The Econometrics Journal*, 2010, *13* (2), 145–176.

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

— **and Bruce A Weinberg**, "Unrequited friendship? How reciprocity mediates adolescent peer effects," *Regional Science and Urban Economics*, 2014, *48*, 144–153.

Liu, Xiaodong, Eleonora Patacchini, and Yves Zenou, "Endogenous peer effects: local aggregate or local average?," *Journal of Economic Behavior & Organization*, 2014, *103*, 39–59.

Manski, Charles F, "Dynamic choice in social settings: Learning from the experiences of others," *Journal of Econometrics*, 1993, *58* (1-2), 121–136.

— , "Identification of endogenous social effects: The reflection problem," *Review of Economic Studies*, 1993, *60* (3), 531–542.

Massen, Jorg JM and Sonja E Koski, "Chimps of a feather sit together: chimpanzee friendships are based on homophily in personality," *Evolution and Human Behavior*, 2014, *35* (1), 1–8.

McEwan, Patrick J, "Peer effects on student achievement: Evidence from Chile," *Economics of education review*, 2003, *22* (2), 131–141.

Moffitt, Robert A et al., "Policy interventions, low-level equilibria, and social interactions," *Social dynamics*, 2001, *4* (45-82), 6–17.

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou, "Mothers, peers, and gender-role identity," *Journal of the European Economic Association*, 2018.

Olsen, Randall J, "A least squares correction for selectivity bias," *Econometrica*, 1980, pp. 1815–1820.

Patacchini, Eleonora and Giuseppe Venanzoni, "Peer effects in the demand for housing quality," *Journal of Urban Economics*, 2014, *83*, 6–17.

_ **and Yves Zenou**, "Racial identity and education in social networks," *Social Networks*, 2016, *44*, 85–94.

_ **, Edoardo Rainone, and Yves Zenou**, "Heterogeneous peer effects in education," *Journal of Economic Behavior & Organization*, 2017, *134*, 190–227.

Porter, Catherine and Danila Serra, "Gender differences in the choice of major: The importance of female role models," *American Economic Journal: Applied Economics*, 2019.

Raftery, Adrian E, Michael A Newton, Jaya M Satagopan, and Pavel N Krivitsky, "Estimating the integrated likelihood via posterior simulation using the harmonic mean identity," in "Bayesian Statistics" Citeseer 2007.

Riley, Emma et al., "Role models in movies: The impact of Queen of Katwe on students educational attainment," Technical Report, Centre for the Study of African Economies, University of Oxford 2017.

Sabia, Joseph J and Daniel I Rees, "Does the number of sex partners affect educational attainment? Evidence from female respondents to the Add Health," *Journal of Population Economics*, 2012, *25* (1), 89–118.

Sacerdote, Bruce, "Peer effects with random assignment: Results for Dartmouth roommates," *Quarterly Journal of Economics*, 2001, *116* (2), 681–704.

Walsh, Patrick, "Are involved parents providing public goods or private goods?," *Public Finance Review*, 2008, *36* (6), 678–705.

Wang, Dan J, Xiaolin Shi, Daniel A McFarland, and Jure Leskovec, "Measurement error in network data: A re-classification," *Social Networks*, 2012, *34* (4), 396–409.

Wilson, William Julius, *The truly disadvantaged: The inner city, the underclass, and public policy*, University of Chicago Press, 2012.

Zimmerman, David J, "Peer effects in academic outcomes: Evidence from a natural experiment," *The Review of Economics and Statistics*, 2003, *85* (1), 9–23.

A Sample Characteristics

Table A1: Appendix: Sample characteristics do not vary after dropping observations

	(1) Traced Sample from Wave I		(2) Only connected students		(3) Network size between 11 and 400	
	mean	sd	mean	sd	mean	sd
College Completion (Wave IV)	0.35	0.48	0.36	0.48	0.36	0.48
Female	0.51	0.50	0.53	0.50	0.53	0.50
Age	14.80	1.77	14.73	1.75	14.66	1.76
Black	0.14	0.34	0.13	0.33	0.13	0.34
Other	0.19	0.39	0.18	0.38	0.17	0.38
Multiple Races	0.01	0.09	0.01	0.09	0.01	0.09
Friend nominations	4.22	2.89	4.95	2.48	4.95	2.49
Family and community characteristics						
Father with college degree	0.24	0.43	0.25	0.44	0.25	0.43
Father as Professional	0.19	0.39	0.20	0.40	0.20	0.40
Mother with college degree	0.25	0.43	0.26	0.44	0.26	0.44
Mother as Professional	0.27	0.44	0.27	0.44	0.28	0.45
Two-Parent Family	0.72	0.45	0.74	0.44	0.74	0.44
Race Dispersion (Block level)	0.22	0.24	0.21	0.24	0.21	0.23
Crime rate (county level)	5215.58	2784.61	5050.58	2739.44	5134.58	2765.12
Median income ($1,000) (block level)	29.94	13.18	30.03	12.98	29.60	13.23
Observations	10,258		8,563		7,399	

Note: This table shows that the sample characteristics do not change much with the two sample selection criterion. Apart from the outcome of interest 'College completion', this table also shows all control variables that are included in the estimation. Especially, 'Family and community characteristics' are used to address neighborhood sorting. The final sample consists of 7,399 students from 116 networks (schools). Cross-sectional weight in Wave 4 applies.

B Running a Linear Probability Model for Interpretation

Table A2: Appendix: Spillover from Peers' Parents Diminishes With Care From Own Father (Linear Probability Model)

	Two-Parent Family		
	(1)	(2)	(3)
VARIABLES	Care from both	Care from mom	Care from dad
CareIndex*(Peers' Paternal Education on Boys)	-0.0972**	-0.0260	-0.102**
	(0.0482)	(0.0502)	(0.0499)
CareIndex*(Peers' Maternal Education on Boys)	0.0318	0.0450	0.0178
	(0.0462)	(0.0487)	(0.0472)
CareIndex*(Peers' Paternal Education on Girls)	-0.00669	0.0350	-0.0137
	(0.0409)	(0.0470)	(0.0425)
CareIndex*(Peers' Maternal Education on Girls)	-0.0216	-0.0104	-0.0229
	(0.0383)	(0.0473)	(0.0394)
CareIndex	0.0586***	0.00998	0.0546***
	(0.0147)	(0.0108)	(0.0151)
Peers' Paternal Education on Boys	0.187***	0.146***	0.189***
	(0.0435)	(0.0385)	(0.0435)
Peers' Maternal Education on Boys	-0.0506	-0.0443	-0.0458
	(0.0409)	(0.0371)	(0.0409)
Peers' Paternal Education on Girls	0.0253	0.0226	0.0282
	(0.0366)	(0.0349)	(0.0367)
Peers' Maternal Education on Girls	0.113***	0.106***	0.114***
	(0.0333)	(0.0324)	(0.0333)
Constant	-0.254	-0.192	-0.248
	(0.324)	(0.324)	(0.325)
Observations	5,453	5,453	5,453
R-squared	0.373	0.371	0.372

Note: Robust standard errors in parentheses; ***, ** and * represent 1%, 5% and 10% significance level respectively. Dependent variable in all regressions is a college completion indicator. All regressions include control variables in Equation 1, grade and school fixed effects.

Four measures about parental cares are obtained from "How close do you feel to your mother/father?" and "How much do you think she/he cares about you?". The responses are recorded in the Wave 1 In-Home survey of AddHealth. 'CareIndex' is obtained by analyzing the correlation matrix of the measures using the principal-factor method. The 'CareIndex' in Coumn 1 is obtained using all the four measures, whereas that in Column 2 and 3 are obtained using mother(father)-specific measures. The sample is constrained to two-parent families (5,453 students) because responses are skipped for students from single parent families.

C Heterogeneity by the Rank of Friend

Table A3: Appendix: The Gender Specific Pattern Concentrates On Close Friends (Break-Down)

	Boy Friends					Girl Friends				
	First	Second	Third	Fourth	Fifth	First	Second	Third	Fourth	Fifth
Peers' Fathers on Males	0.165**	0.227***	0.0498	0.0895	0.122	0.146*	0.0237	-0.115	0.0873	0.169
	(0.0760)	(0.0823)	(0.0879)	(0.100)	(0.117)	(0.0844)	(0.0923)	(0.102)	(0.112)	(0.143)
Peers' Mothers on Males	-0.0794	-0.0879	0.0576	-0.0687	0.187	0.0553	0.00134	-0.0172	0.0106	-0.0976
	(0.0759)	(0.0835)	(0.0858)	(0.0983)	(0.114)	(0.0834)	(0.0895)	(0.0999)	(0.114)	(0.138)
Peers' Fathers on Females	0.0285	-0.0823	0.0249	-0.0992	-0.165	0.0313	0.0539	0.143*	0.0890	-0.0427
	(0.0789)	(0.0828)	(0.0920)	(0.111)	(0.145)	(0.0685)	(0.0709)	(0.0757)	(0.0803)	(0.100)
Peers' Mothers on Females	-0.0182	0.145*	0.261***	0.187*	0.0920	0.138**	0.112	-0.0472	0.0918	0.0452
	(0.0727)	(0.0783)	(0.0897)	(0.106)	(0.144)	(0.0656)	(0.0688)	(0.0731)	(0.0777)	(0.0947)

Dependent variable is a college completion indicator. A selection-corrected approach with two-dimensional error-correction terms are estimated. The MCMC estimation runs for 150,000 iterations with the first 40,000 iterations as burn-in. Standard deviation of the 110,000 posterior draws is in the corresponding parenthesis. Hypothesis testing is based on frequentist's approach, where ***, ** and * represent 1%, 5% and 10% significance level respectively. All coefficients are estimated simultaneously in one regression. The regression includes control variables in Equation 1, grade and school fixed effects.

D MCMC Algorithm

Define Y be the outcome variable, X and Z be the observed characteristics in outcome and network equation respectively, and ξ be the d-dimensional latent factors. D_i represents all observed links of student i. Let also Θ be the set of all parameters. The likelihood function for each school g is then:

$$\begin{aligned} L(Y_s, D_s|X_s, Z_s, \xi_s; \Theta) &= \int_{\xi_s} P(Y_s^*, D_s|X_s, Z_s, \xi_s; \Theta) f(\xi_s) d\xi_s \\ &= \int_{\xi_s} P(Y_s^*|D_s, X_s, \xi_s; \Theta, \alpha_s) P(D_s|Z_s, \xi_s; \Theta) f(\xi_s) d\xi_s \quad (7) \end{aligned}$$

The estimation of the above likelihood function procedure closely follows (Hsieh and Lee, 2016), which Metropolis-Hasting (M-H) algorithm is incorporated in Gibbs sampling.

Let y_i^* be agent i's latent variable of the outcome equation and follows normal distribution. The subscript for each school s is dropped unless specified. For clarity, let $\boldsymbol{\beta} = \{\beta, \delta, \phi\}$ and $\theta = \{\boldsymbol{\beta}, \alpha, \rho, \gamma\}$.

The prior distributions of the parameters and the unobserved latent factors are defined as:

$$\begin{aligned} \xi_i &\sim N_d(0, I_d) \\ \gamma &\sim N_q(\gamma_0, \Gamma_0) \\ \boldsymbol{\beta} &\sim N_k(\boldsymbol{\beta}_0, B_0) \\ \rho_d &\sim N_d(\rho_0, \sigma_{d0}) \\ \alpha_g &\sim N(a_0, A_0) \end{aligned}$$

For each iteration, we draw a new set of values for the parameters according to the following procedures:

Latent variable: The full conditional of $y^*|\theta, Z, Y, X, W$ is a truncated normal distribution, that is

$$P(y^{*(t)}|\theta^{(t-1)}, \xi^{(t-1)}, Y, X, W) = \mathbf{1}(Y_i = 1)\mathbf{1}(y_i^* > 0) + \mathbf{1}(Y_i = 0)\mathbf{1}(y_i^* \leq 0)$$

Sample $\{y_i^{*(t)}\}$ from the aforementioned posterior distribution.

Unobserved ξ: Sample $\{\xi_i^{(t)}\}$ from $P(\xi^{(t)}|y^{*(t)}, \theta^{(t-1)}, Y, X, W)$ with M-H, where

$$P(\xi^{(t)}|y^{*(t)}, \theta^{(t-1)}, Y, X, W) \propto N(\xi; 0, I) P(y^*|W, \xi; \theta^{(t-1)}) P(W|\xi, \gamma^{(t-1)})$$

This procedure is repeated for each network independently. Adaptive updating is employed to achieve the optimal acceptance rate between 20% and 30%.

Link formation: Sample γ from $P(\gamma|W, \{\xi^{(t)}\})$ with M-H, where

$$P(\gamma|W, \{\xi^{(t)}\}) \propto N_{q+2}(\gamma; \gamma_0, G_0)P(W|\xi_i^{(t)}, \xi_j^{(t)}, \gamma)$$

Outcome parameters: Sample $\boldsymbol{\beta}$ from $P(\boldsymbol{\beta}|y^{*(t)}, W, X, \xi^{(t)}; \rho^{(t-1)}, \alpha^{(t-1)})$, where

$$\begin{aligned} P(\boldsymbol{\beta}|y^{*(t)}, W, \xi^{(t)}; \rho^{(t-1)}, \alpha^{(t-1)}) &\propto N(\theta_0, Q_0)P(y^*|W, X, \xi^{(t)}; \rho^{(t-1)}, \alpha^{(t-1)}, \boldsymbol{\beta}) \\ &\propto N_k(\mathbf{M}, \mathbf{B}) \end{aligned}$$

with $\mathbf{M} = \mathbf{B}(Q_0^{-1}\theta_0 + X'(y^* - \xi\rho - l\alpha))$ and $\mathbf{B} = (B_0^{-1} + X'X)^{-1}$.

Error correction: Sample ρ from $P(\rho|y^{*(t)}, W, X, \xi^{(t)}; \boldsymbol{\beta}^{(t)}, \alpha^{(t-1)})$ with M-H, where

$$P(\rho|y^{*(t)}, W, X, \xi^{(t)}; \boldsymbol{\beta}^{(t)}, \alpha^{(t-1)}) \propto N(\rho_0, \sigma_0)P(y^*|W, X, \xi^{(t)}; \boldsymbol{\beta}^{(t)}, \alpha^{(t-1)}, \rho)$$

Group effects: Sample α_g from $P(\alpha_g|y_g^{*(t)}, W_g, X, \xi_g^{(t)}\}; \beta^{(t)}, \rho^{(t)})$, where

$$\begin{aligned} P(\alpha_g|y_g^{*(t)}, W_g, X, \xi_g^{(t)}\}; \boldsymbol{\beta}^{(t)}, \rho^{(t)}) &\propto N(\alpha_0, A_0)P(y_g^*|W_g, X_g, \xi_g^{(t)}; \boldsymbol{\beta}^{(t)}, \rho^{(t)}, \alpha_g) \\ &\propto N(\hat{\alpha}_g, R_g) \end{aligned}$$

with $\hat{\alpha}_g = R_g(A_0^{-1}\alpha_0 + l_g'(y_g^* - X_g\beta - \xi_g\rho))$ and $R_g = (A_0^{-1} + l_g'l_g)^{-1}$

Appendix: Convergence Diagnosis

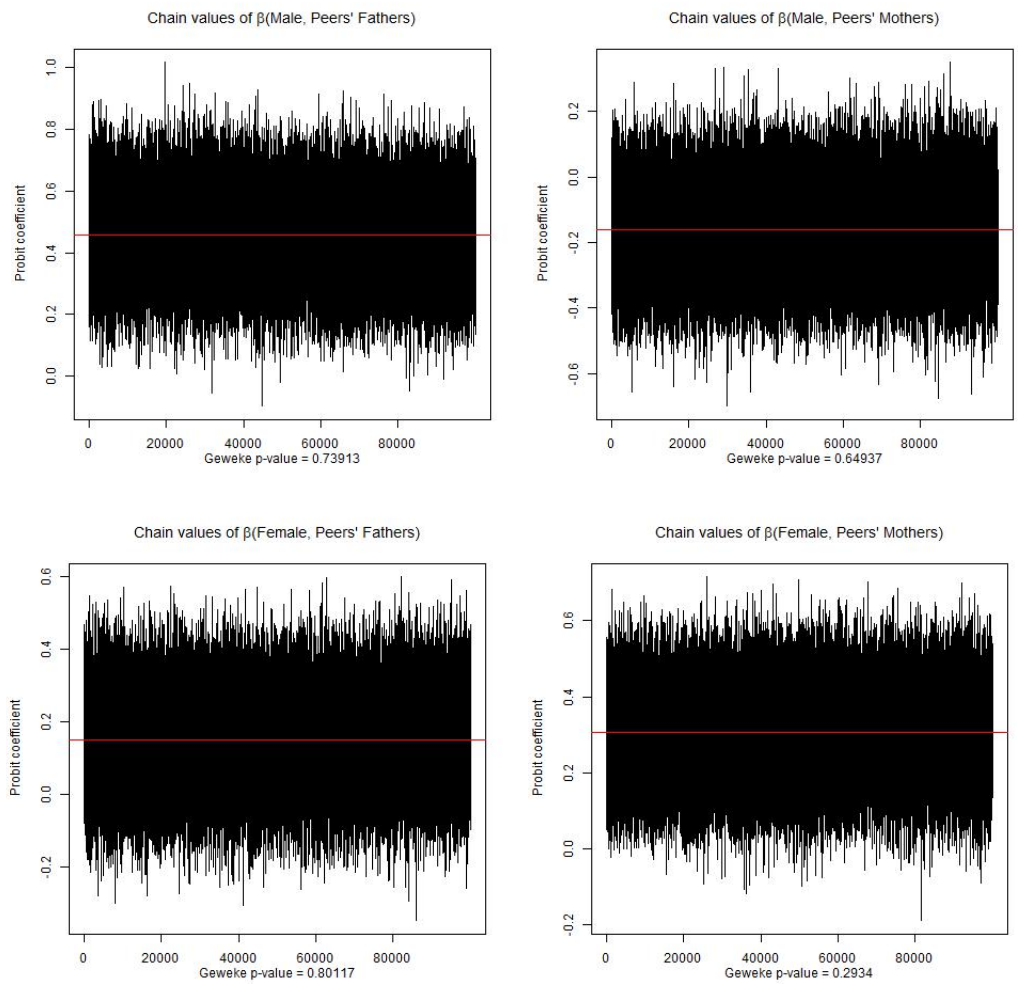

Figure A1: The figures show the chain values of the four variables of interest. Convergence is confirmed by Geweke (1992)'s diagnostic that mimics a simple two-sample test of means between the first 10% and the last 50% of the chain values.

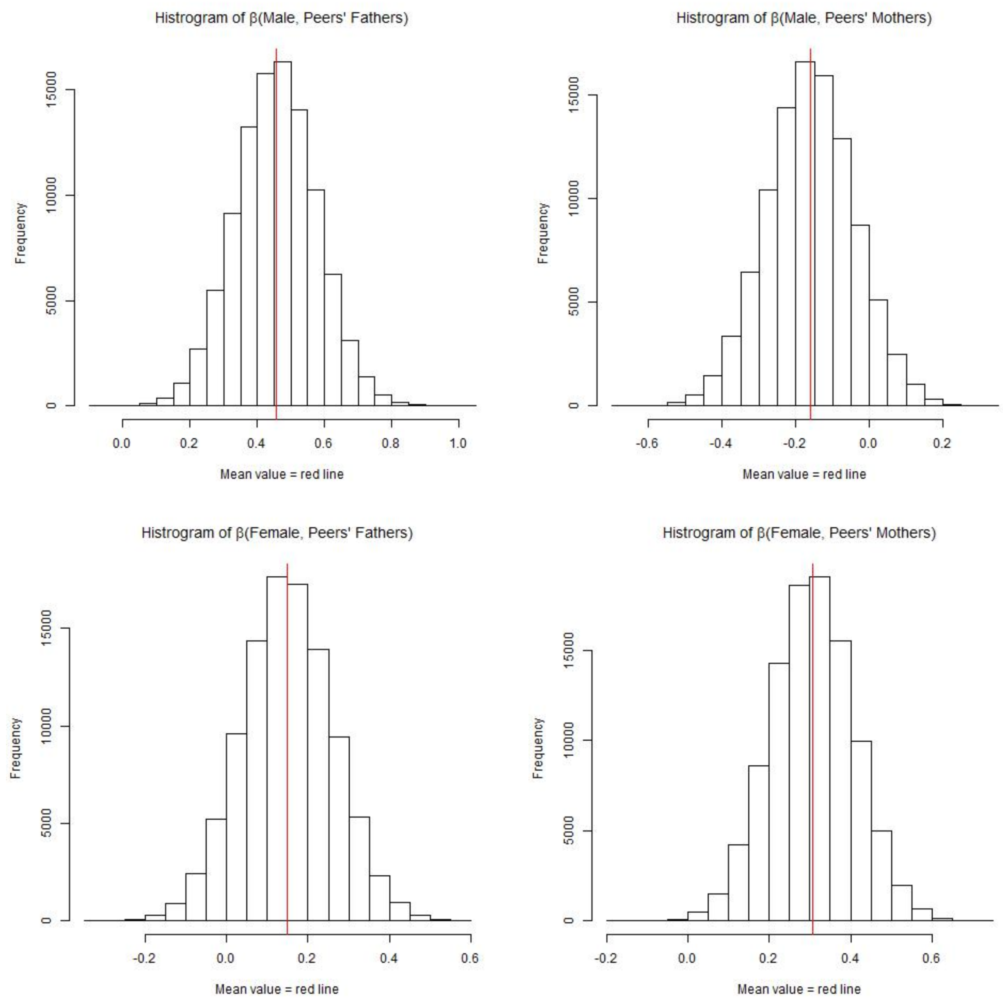

Figure A2: The figures show the histograms of the draws of the four variables.