HCEO WORKING PAPER SERIES

Peers' Parents and Educational Attainment

Bobby Chung

Working Paper

2018-086

11/2018

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 *

Bobby W. Chung †

November 3, 2018

Abstract

This paper contributes to the discussion on childhood exposure 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 and find that the spillover from peers' parents of the same gender operates independently of peer effects. The effects are robust to addressing friendship selection. The same gender pattern suggests either the transmission of gender-specific information or the presence of a role model effect. Furthermore, the same gender spillover is significant only for students from lower-educated families. A student whose father is absent or less caring also experiences significant influence from peers' fathers. The heterogeneity by own family background indicates the influences from parental and non-parental adults are substitutes.

JEL Classification: C11, D91, I24, J10

Keywords: Peers' Parents, Social Interaction, College Attainment, Childhood Exposure

---

*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.

†John E. Walker Department of Economics, Clemson University, 424 Sirrine Hall, Clemson, SC 29634. Email: wingyic@clemson.edu. Phone: 1-864-633-8236. 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 Scott Barkowski, Steven Durlauf, Alex Eble, Paul Goldsmith-Pinkham, Amanda Griffith, Majid Hashemi, Chih-Sheng Hsieh, Tommy Leung, Elijah Neilson, Benjamin Posmanick, Curtis Simon, Senad Sinanovic, Kenneth Whaley, and Dan Wood. Special thanks to the seminar participants at Clemson, UChicago, Wake Forest, and the Fourth Conference on Network Science and Economics. All remaining errors are mine.

1 Introduction

Education provides labor market signals, leads to the accumulation of human capital, and affects important labor market outcomes. A key focus of the literature on education is to understand why early life experience affects educational outcomes (Heckman and Mosso, 2014). One explanation deals with the inter-generational transmission of human capital from own parents. Another growing strand in the literature emphasizes peer effects among close friends. Following logically the idea of these two intellectual streams, this paper empirically shows that the educational outcomes of young people are also influenced by their peers' parents.

There are at least three reasons to believe that peers' parents can affect the educational outcomes of a child. First, there exists a direct influence among peers (Sacerdote, 2001). Children learn from their own parents to behave in certain ways and create pressure to conform among their peers. Second, parents communicate the benefits of educational achievement to their child's friends either directly through interaction or indirectly through the social network (Granovetter, 1973, 1983; Jackson, 2010). The theory about the spread of information among social ties is also well-supported by empirical work.[1] Third, young people make choices based on the outcomes realized by a set of 'role models' (Manski, 1993a). Bisin and Verdier (2001) construct a similar theoretical framework that cultural transmission works through horizontal (oblique) socialization by non-parental adults.[2]

To measure the influence of peers' parents on the educational decision of a child, I investigate the extent to which the educational background of peers' parents is related to a child's 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, it also contains the details

[1] Previous empirical studies investigate how the spread of information affects criminal activity (Patacchini and Zenou, 2008), job search (Wahba and Zenou, 2005), career choice (Tümen, 2017), and the drop-out decision of high-school students (Coleman, 1988)

[2] Subsequent empirical papers also show the importance of learning by observing and imitating non-parental adults on child development (Borjas, 1995; Chetty et al., 2018).

of friendship networks in the same school and the educational background of parents. I find evidence that the educational background of peers' parents does influence a child's college attainment. Moreover, the mechanisms are based on the gender of both peers' parents and the child. Whereas the inclusion of direct peer effects and neighborhood characteristics negate the effects 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 also comparable to the influence from stronger social ties. On average, having one peer's parent of the same gender who graduates from college increases the likelihood of completing college for boys and girls by 2.36 and 1.58 percentage points respectively. The same gender spillover on boys is as important as one-third of the effect of having a father who is a college graduate, whereas the same gender spillover on girls is as important as one-sixth of the effect of having a mother who is a college graduate. The same-gender influences on both boys and girls are also comparable to half of the effect from a one standard deviation increase in average GPA of peers.

One might worry that students do not choose their friends randomly and the positive sorting among students could explain the results that I document. To address selection due to unobserved heterogeneity, I estimate a selection-corrected model that characterizes the network formation process based on homophily (individual bonding based on similarity) (Hsieh and Lee, 2016; Hsieh and Lin, 2017). In the formation process, the distance of latent factors captures homophily based on unobserved attributes. The latent factors then enter into the outcome equation as control functions. Conditional on the selection process specified in the model, the peers' characteristics and the disturbance in the outcome equation are orthogonal (Brock and Durlauf, 2001; Blume et al., 2010). This approach treats the selection problem as omitted variable bias and combines the spirits of the Heckman two-step procedure (Heckman, 1976, 1979) and the control function approach.[3] The same-gender effects remain robust to addressing friendship selection, and the corrections in upward bias occur mainly in

[3]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).

the direct peer effects.

There are two explanations for why the same gender spillover from peers' parents operates independently of peer effects. First, there has been empirical evidence on the positive impacts of having gender-specific adult role models on education-related outcomes (Bettinger and Long, 2005; Dee, 2007; Griffith, 2014; Eble and Hu, 2017). The influence can involve human capital transfer, sharing information or changing preference through direct contact (Chung, 2000). The role model effect also reconciles the different results I find for own parents, that both males and females experience a similar magnitude of the influence from own father and mother, because one's own parents can communicate the information or transfer human capital through daily contact. Second, the spread of information is gender-specific (Akerlof and Kranton, 2000, 2002). In the model of gender identity, utility-maximizing individuals care about self-image, and 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. In the current context, the assimilation of the 'ideal behaviors' can work through direct contacts or indirectly through the social network.

I further find evidence that influences from parental and non-parental adults are substitutes. 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] Third, the same gender spillover on males diminishes with the intimacy with own father. Changing from having the least caring to the most caring father completely offsets the same gender spillover on males. When own parents are lower-educated, less caring, or even absent, the influences from well-educated non-parental adults become more important (Bisin and Verdier, 2001).

The evidence on peers' parents here advances our understanding of social interaction by empirically showing that the composition of a network affects the type of information being transmitted. The primary focus of the literature on social influence is on the peer effects

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

among strong social ties (Sacerdote, 2001; Zimmerman, 2003; Patacchini et al., 2017). As Manski (1993b) points out, the pre-determined characteristics of a group are as important as the contemporaneous spillover in driving social influences, but with different policy implications. For example, in enhancing the effectiveness of a tutoring program, the existence of spillover effects is crucial. A program does not need full coverage to benefit all students because there will be spillovers through the social multiplier. In contrast, the study on the family background of peers can help answer different types of questions such as the socioeconomic composition of a neighborhood or classroom.

The socioeconomic composition of a peer group is also a crucial aspect of social mobility. There have been empirical works showing that school and neighborhood composition affect child development (Katz et al., 2001; Hanushek et al., 2009; Chetty et al., 2016a). One important source of the effect is the exposure to surrounding environments (Chetty et al., 2016b; Chyn, 2017; Chetty and Hendren, 2018). Although this paper does not speak to a particular policy, composition of social ties can be one essential element of the 'exposure effect'. The presence of well-educated adults could benefit children who are from the deprived background the most.

2 Literature on Peer Effect and Peers' Parent

The seminal work by Manski (1993b) pioneers the distinction between contemporaneous spillover among peers and the effect of peers' characteristics (contextual effect). Subsequent works on social interaction pay most of the attentions to identify the contemporaneous spillover using econometric techniques (Lee, 2007; Bramoullé et al., 2009; Lee et al., 2010; Lin, 2010), or random assignment of peers (Sacerdote, 2001; Zimmerman, 2003).

This paper is more related to a few studies that look at contextual effects being caused by the family background of peers. Without data on individual social network, earlier attempts to estimate the spillover from peers' parents assume students are connected to and affected equally by everyone in the same network. Peers either refer to classmates, grade-mates or schoolmates. 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. 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. Olivetti et al. (2015) also analyze the AddHealth data and find significant effects from the weekly hours worked by mothers of grade-mates on labor force participation of females.[5]

Further attempts analyzing friendship network in AddHealth data define the educational background of peers' parents differently, and therefore find mixed results. For example, Hsieh and Lee (2016) and Hsieh and Lin (2017) estimate spillovers *among* peers. They include only the education and job status of peers' mothers as control variables and find insignificant effects. In contrast, Patacchini and Zenou (2016) investigate the effects from having same-race friends. They control for education status of either peers' fathers or peers' mothers, depending on who is the interviewee in the survey, and find positive effects.

This paper complements the above studies in three aspects. First, I examine the friendship networks. Instead of assuming a completed network, students have their own peer groups in the same network. Second, I differentiate the effects of peers' fathers from that of peers' mother. Third, I also relax the assumption that the spillover from peers' parents on educational outcome is the same for male and female students.

3 Methodology

3.1 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 the purpose of this paper

[5]In their robustness section, they also analyze the friendship networks and find a similar result.

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.[6] 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 student's network. 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. In Section 5, this assumption is re-examined. A network is defined as a school.

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).[7] While the literature on contemporaneous effects using this dataset is extensive, only a few studies pay attention to long-term outcomes (Olivetti et al., 2015; Patacchini et al., 2017).

Following the literature, networks which are too small ($<$ 11 students) and too big ($>$ 400 students) are not included in the main analysis.[8] Students with no friend nominations are

[6]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).

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

[8]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. To check sensitivity, the estimation of simple Probit with all networks in Column (2) of Table 1 indeed gives the same qualitative results. For clarity, results are available upon request.

dropped. In total, 28% of the observations in the original sample are discarded. Table 1 shows that the summary statistics of the variables 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.

Two sets of variables are important to disentangle the mechanisms through which the spillover from indirect social ties operate. First, direct peer effects are measured by the peers' characteristics including age, race, gender, as well as GPA which is the average of English, Mathematics, History and Science. Second, the variables listed under 'Family and community characteristics' in Table 1 are used to proxy for the choice of parents. These variables can capture their choices on neighborhood and schools that cannot be addressed by network fixed effects.

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 father and mother of friends. Table 2 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 a degree, compared to 39% for males with a degree. The difference between two is also statistically significant. This pattern applies also the average college attainment of peers' mothers. Although there exists a positive relationship between the educational background of peers' parents and college completion of a student, the selection issue is still a concern in causal inference. Also, students benefit from peers' family background may simply due to the interactions with better peers. The next two sections discuss the causal estimation and the isolation of peer effects in details.

3.2 Sociomatrix and Baseline Model

To formalize the idea of social interactions 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 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}.[9] 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:

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

$FATHER_{js}$ 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 male and female students. For example, $\beta_{male,FATHER}$ is the effects on male students from having more peers' fathers who are college graduates. X_{isg} is a vector of i's

[9]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.

characteristics which are shown in Table 1. 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 the effects of peers' quality. Because the number of friends varies, out-degree is also included as a control variable.

Using the terminology by Manski (1993b), the coefficients of interest (β) in Equation 1 are the estimates of contextual effects, i.e. the effects of the pre-determined characteristics of peers. Since the focus of this paper is not the spillover *among* peers, the average GPA of peers is contained in $W_s X_{js}$ and enters the regression as a control variable instead of being the main variable of interest. This also gives β the interpretation of the spillover from peers' parents that operates independently of the direct peer effects. One alternative to isolate the influences of peers' parents from peers is controlling for the average college attainment rate of peers (Patacchini et al., 2017). However, as discussed in the data section, only a subset of students in the friendship network in Wave I is traced through Wave IV. To preserve the network structure, average GPA of peers is a reasonable substitute.

3.3 Selection-Corrected Model

The identifications of β in Equation 1 rely on the independence of the sociomatrix D_s (more specifically the $d_{ij,s}$ entries) and u_{isg}. However, this does not hold in many circumstances as relationship sorting is well-documented (Jackson, 2010; Carrell et al., 2013). Although including network fixed effects and controlling for family backgrounds can mitigate the problem from neighborhood sorting, the list of control variables can hardly be exhaustive in dealing with relationship sorting based on unobserved characteristics within a network (Manski, 1993b; Bramoullé et al., 2009). In this paper, the selection issue is handled as omitted variable bias. That is, Equation 1 fails to incorporate common factors that simultaneously determine the outcome and link decisions.

To formalize this strategy, I first characterize friendship formation using the latent space model (Hoff et al., 2002). This model of endogenous link formation is based on a social phenomenon called 'homophily', a term is coined in Lazarsfeld et al. (1954). In models of homophily, two agents are more likely to have a tie if they share similar characteristics. These shared traits can be observed characteristics such as gender and age, or unobserved traits such as personality. The distances of the characteristics between two agents will be the key explanatory variables in the endogenous friendship formation.

The outcome variable in this model, $d_{ij,s}$, equals 1 if i sends a link to j in a network, and the network model is in a logit form.[10] The probability of each link equals 0 or 1 is:

$$Pr(d_{ij,s}|\psi_{ij,s}) = \left(\frac{1}{1+exp(-\psi_{ij,s})}\right)^{d_{ij,s}} \left(\frac{exp(-\psi_{ij,s})}{1+exp(-\psi_{ij,s})}\right)^{1-d_{ij,s}} \tag{2}$$

where

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

The above structural specification closely follows that of (Hsieh and Lee, 2016; Hsieh and Lin, 2017). $\mathbf{X}_i$ and $\mathbf{X}_j$ contain i's and j's age respectively. The dyad-specific (absolute distance of gender, age and grade) observed characteristics explain the link decision based on homophily and serve as exclusion restrictions. $e_{ik,s}$ refers to the k-th unobserved latent factor (such as attitude or habit) that matters in the link decision, and $\bar{d}$ represents the dimension of the latent factors which represent different unobserved traits. $|e_{ik,s} - e_{jk,s}|$ refers to the absolute distance of the latent factor between i and j. γ_k and the coefficients of the dyad-specific variables are expected to be negative based on the idea of homophily. The key of this model is the vector of $\bar{d}$-dimensional latent factors $\xi_{i,s} = (e_{i1,s}, ..., e_{i\bar{d},s})$ which provides a linkage

[10]Another possible parametric assumption is using a probit link (Chan and Lam, 2014).

between network formations and the outcome equation.

To incorporate the latent factors from link formations which are not observed, the outcome equation in Equation 1 is augmented using the approach of Albert and Chib (1993). Assume y^*_{isg} represents the latent variable underlying the decision of college attainment. The augmented version of Equation 1 becomes:

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

$$\begin{aligned} y^*_{isg} = {} & \beta_{male,FATHER} W_s^{male} FATHER_s + \beta_{male,MOTHER} W_s^{male} MOTHER_s \\ & \beta_{female,FATHER} W_s^{female} FATHER_s + \beta_{female,MOTHER} W_s^{female} MOTHER_s \\ & + W_s X_s \delta + X_{isg}\phi + \alpha_g + \alpha_s + \sum_{k=1}^{\bar{d}} \rho_k e_{ik,sg} + \epsilon_{isg} \end{aligned} \quad (3)$$

and ϵ_{isg} follows standard normal distribution. The breakdown of u_{isg} in Equation 1 into $\sum_{k=1}^{\bar{d}} \rho_k e_{ik,sg} + \epsilon_{isg}$ requires the parametric assumptions that $E[u_{isg}|\xi_{i,s}]$ being linear in $\xi_{i,s}$.[11] The $\bar{d}$-dimensional latent factors enter Equation 3 as control functions, and thus the sociomatrices are conditionally independent of ϵ_{isg}. The intuition is that if the similarity in unobserved attributes, such as beliefs and habits, significantly affect whether i and j bond together, and if these attributes significantly determine the outcome ($\rho > 0$), failure to account for it could result in biased estimates of β. One important issue regarding identification. Without excluded variables in $\psi_{ij,s}$ in Equation 2, the identification of β relies on the non-linearity of $e_{i,s}$, and this can cause imprecise estimates (Brock and Durlauf, 2003). Therefore, dyad-specific characteristics in Equation 2 are used as exclusion restrictions. The assumption is that similarities in observed characteristics, controlling for the characteristics

[11] 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).

in the outcome equation, only affect the outcome y^*_{isg} through changing the likelihood of forming the link. The threat to this identifying assumption is that the characteristics are determined after friendships being formed such as common club activities.[12] Therefore, the observed characteristics are all pre-determined.[13]

In the above structural model from Equation 2 and Equation 3, 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 D_s for the estimation is:

$$\begin{aligned} L(Y^*_s, D_s|X_s, Z_s, \xi_s; \Theta) &= \int_{\xi_s} P(Y^*_s, D_s|X_s, Z_s, \xi_s; \Theta) f(\xi_s) d\xi_s \\ &= \int_{\xi_s} P(Y^*_s|D_s, X_s, \xi_s; \Theta, \alpha_s) P(D_s|Z_s, \xi_s; \Theta) f(\xi_s) d\xi_s \end{aligned} \quad (4)$$

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. To circumvent this difficulty, I estimate link formations and the selection-corrected outcome equation simultaneously using Bayesian method (mixing Metropolis-Hastings (M-H) and Gibbs sampler).[14] Details of MCMC algorithm are described in Appendix A. 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-

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

[13] 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.

[14] As noted by Hsieh and Lee (2016), this approach is essentially full information maximum likelihood.

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.

Besides the parametric assumptions, the defect of the aforementioned approach is computational burden. An alternative approach to address selection bias in non-experimental analysis is to instrument for the average characteristics of peers Bramoullé et al. (2009).[15] However, link formations (and thus the adjacency matrix) are still assumed to be independent of individual choices. Also, monotonicity assumption in IV estimation may be violated when behavioral responses of group sorting are taken into account (Echenique and Fryer Jr, 2007).

4 Results

4.1 The Role of Peer Effects and Parental Choice

First, I estimate a simple probit model (without the network formation model) and check out the role of direct peer effects and the choice by parents on neighborhood in explaining the spillover from peers' parents.

Following previous studies that also analyze AddHealth data, 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 3, 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

[15]Sacerdote (2001) and Zimmerman (2003) are two seminal works that exploit random assignments of peers. Even in these circumstances, interactions still depend on agents' choices.

Table 2. An interesting observation is that the size of the effects from friends' fathers are twice as large as that from friends' mothers even though both are estimated with similar precision. The difference is also statistically significant (f-stat equals 15.19). The inclusion of own characteristics (GPA, age, race, and gender) does not cause a big change in the magnitude as shown in Column (2). However, when own family background, neighborhood characteristics, and fixed effects are included, the size of the estimates for peers' fathers and peers' mothers are reduced by 50% and 67% as shown from Column (3) to (5).[16] This shows the importance of sorting based on socioeconomic status. In Column (6), I estimate the fully saturated model that takes direct peer effects into account by including peer characteristics (including GPA, race, gender, age and single parent status). Both the effects from friends' fathers and friends' mothers are further reduced by 41% and 37%, and only the former remains statistically significant. Throughout the exercise, the size of the influence from peers' fathers is mostly statistically larger than that from peers mothers, and the difference is explained away by direct peer effects.

The above estimates of the average effects across gender are consistent with the findings in previous studies which also analyze AddHealth data. Hsieh and Lee (2016) treat the family background of friends as control variables using 'the proportion of friends' mothers who graduate from high-school' and the proportion of friends' mothers who work as professionals', and find insignificant effects.[17] Patacchini and Zenou (2016) indeed find significant and positive effects from parental education of friends. However, they define 'parent' as the interviewee in the In-Home survey and therefore can either be father or mother. Therefore, their estimate is essentially mixing the effects from friends' fathers and friends' mothers.

I repeat the exercise above to examine the role of direct peer effects and parents' choices on neighborhood in explaining the heterogeneous effects across gender from peers' parents in

[16] 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 Section4.2 because a continuous latent variable is assumed Albert and Chib (1993).

[17] 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 4. 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 above, the influence from peers' fathers is larger than that from peers' mothers. However, the gender-specific pattern on males starts to emerge because only the difference on male students are statistically significant (f-stat: 21.66, as opposed to 2.03 for female students). Again, adding own characteristics as shown 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 males in which the magnitude drops by more than a half and it becomes statistically insignificant. Although the size of same-gender spillover (for both male and female) and the effects from peers' fathers on females become smaller, the magnitudes 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 females drops by 59% and becomes insignificant. The same-gender spillovers for male and female students also decrease by 29% and 15%, but they remain statistically significant.

The decomposition exercise tells the importance of direct peer effects and the choice of parents. In particular, comparing the fully-saturated model in Column (6) and the simple model in Column (2) of Table 4, the two factors explain about half of the same-gender spillover and completely explain away the opposite-gender spillover. What remain statistically significant, after considering peer effects and parents' choices, are the same-gender spillover and the difference of the effects between peers' fathers and peers' mothers on male students.

Table 5 compares the marginal effects from peers, peers' parents, and own parents. 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 male students, the effects from having one more peers' father with college degree are one-third of the effects from own parents. For female students, the effects from having one more peers' mother with college degree are one-sixth of the effects from own parents. 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 one standard deviation in average peers' GPA.

4.2 Selection-Correction

In the fully-saturated model above, friendship formation is assumed to be exogenous. To address the selection issues, I estimate the network formation model (Equation 2) and the outcome (Equation 3) jointly. Following Hsieh and Lee (2016), I run 150,000 iterations with the first 40,000 as burn-in. The point estimates reported in Table 7 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 Appendix B.

For consistency, I still present the result using Bayesian method for the fully-saturated model to obtain AICM in Column (1) of Table 7. The estimates are essentially the same as that using classical approach in Table 4. 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 7 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 four variables decrease only a little when compared to that in Column (1) and the gender-specific pattern remains robust. Also, the drop in the magnitudes is not statistically significant. However, this does not mean the selection-correction method does not function well. I also present the coefficient for the average GPA of peers. Compared to the estimate in Column (1), the magnitude in Column (3) drops by more than one-third of the standard deviation of the posterior draws.

5 Further Analysis

5.1 Robustness Check

In this section, I check the sensitivity of the above results by varying the definition of friendships. In the main analysis, a link is directed without consensus. That is, the sample pools both reciprocal and non-reciprocal links. The defect of including non-reciprocal links is that interaction may be less close. However, when non-reciprocal links are taken away, there are 6,738 observations remain with only 40% of the links preserved. Therefore, there are concerns for measurement errors in both methods. Including non-reciprocal friendships may fail to reject false links, whereas omitting them may drop the true links. When the inaccuracies are due to random factors such as the survey design, the measurement error is classical which induces attenuation bias. The same gender estimates will be the lower bounds of the true effect, while the opposite-gender spillover may be falsely rejected. In contrast, when the inaccuracies are systematically correspond to the characteristics of respondents, for example smarter students are more honest on their relationships, the estimates of peer effects will then be biased.

To strike the balance in defining friendship, I adopt flexible weighting for non-reciprocal links. That is, I define a weighting factor α varying from 0 to 1. When α approaches 0, less weights are assigned to non-reciprocal links. When $\alpha = 0$, all links are reciprocal friendships, whereas $\alpha = 1$ resembles the weight used in the main analysis where reciprocal and non-reciprocal friends are equally weighted. On one hand, I could preserve the links that are reported. On the other hand, we could track the changes of the results when the weight varies. The four plots in Figure 1 show the changes of the main coefficients when adjusting the weight (α) on non-reciprocal friendships. The estimates are indeed stable across the weighting methods. One exception is the coefficient of the effects of peers' father on males (Top-left panel). When only reciprocal links are counted ($\alpha = 0$), the spillover become small and insignificant. One explanation is that we do lose important information when dropping the non-reciprocal links.

5.2 Substitution Between Parental and Non-Parental Adults

Whereas I found gender-specific spillover from friends' parents, in Table 5 of Section 4, I show that the influences from own parents are gender-neutral. That is, both mother and father exert significant impacts on their children, regardless of whether they are boys or girls. A natural question to ask is: are the influences from peers' parents the same for students from different family backgrounds? 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.[18]

In Table 7, the effects on students across the six family status are reported. The estimation is done by interacting the four variables of interest with indicators of the six family status in one regression to capture the heterogeneous impacts. A linear probability model is employed for the sake of interpretation. For the same-gender spillover on male students, the effects concentrate on students who come from two-parent families with one or none of them graduates from college. Conditional on having 5 friends (the sample average), having one college-grad peers' father increases the likelihood of completing college by 4.1 and 3.3 percentage points which are higher than the average normalized magnitudes in Table 5. The effect on students from a single-mother family is slightly smaller and is marginally significant.

[18]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).

For the same-gender spillover on females, the effect is significant for females only if one of their parents completed college. Again, conditional on having 5 friends, having one college-grad peers' mother increases the likelihood of completing college by 4.4 percentage points for this category of girls.

Two observations may suggest that influences from parental and non-parental adults are substitutes. First, peers' fathers are influential on students from single-mother family. Especially for female students, this is the only family category that I find significant opposite-gender spillover, and the size of the effect is significantly different from that on students from a two-parent family. Second, students from well-educated family (both parents are college grads) do not experience significant effects from peers' parents, regardless of their gender. The size of the effect is smaller than those found in lower-educated households, although the difference in magnitudes is not statistically significant.

To test the hypothesis that there exists substitution between parental and non-parental adults, I further look at whether the magnitude of the spillover from peers' parents diminishes with the intimacy with own parents. AddHealth provides a unique opportunity to explore this question. 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 8 show the summary statistics for the four survey questions. Students usually have a higher rating on mothers than on fathers. 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 results are reported in Table 9. 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. More surprisingly, 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 8). 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. Again, 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.

The exercises in this subsection do suggest that the influences from parental and non-parental adults are substitutes. Moreover, the presence of or the care from the father is crucial in affecting the magnitude of the spillover from peers' parents.

5.3 Other Outcomes

Hsieh and Lin (2017) estimated the effects of peers' parents for other short-term outcomes. They find the educational background of peers' mothers has marginally significant effects (at 10% level) on GPA and no effects on smoking behaviors. Our results differ can be because of their assumption that the effects of peers' parents are homogeneous across gender. To compare my finding with theirs, the same specification in the main analysis, namely differentiating heterogeneous influences from peers' parents by gender, is adopted to analyze the effects on GPA and smoking behaviors. The results are presented from Column (1) to (4) of Table 11.

Column (1) presents the result with the assumption of homogeneous effects across boys and girls. The effects of peers' fathers and peers' mothers on GPA are one-half and one-fourth of a one standard deviation increases in the average GPA of peers. While the effects of peers' fathers are significant at 5% level, that of peers' mothers are marginally significant at 10% level. This result is similar with Hsieh and Lin (2017). Interestingly, there again exists a gender-specific pattern when I examine heterogeneous impacts across gender in Column (2). Also, the difference between the same-gender and opposite gender effects is statistically significant for male students. These two patterns coincide with what I found for college attainment. The smoking outcome does not exhibit the same pattern, however. Neither peers' fathers nor peers' mothers is influential in reducing smoking behaviors.

The above analysis may suggest that reducing incapacitating behaviors is not a mechanism through which peers' parents affect children. To check the sensitivity of my analysis, in addition to smoking behaviors, I also include two other incapacitating behaviors, namely drinking and dangerous behaviors.[19] Not surprisingly, in Table 10, we do observe a strong correlation between GPA and college attainment that the correlation coefficient equals 0.409. This can serve as a benchmark, and the correlation between college attainment and the three incapacitating behaviors are relatively weak. Therefore, regressing these behavioral outcomes on the average college attainment of peers' parents can test the hypothesis that behavioral change is one mechanism.

Column (5) to (8) of Table 11 show that neither peers' fathers nor peers' mothers is influential in reducing drinking and dangerous behaviors. However, this does not mean social influence does not exist. Together with smoking outcome, all the three incapacitating behaviors do exhibit significant spillover effects among peers. For example, in Column (3), if all non-smoking peers smoke, an individual is on average 40% more likely to smoke.

The exercise here shows that the spillover from peers' parents does show up in both short-term and long-term educational outcomes. However, the effects do not work through reducing incapacitating behaviors.

[19] All the variables in this subsection are obtained in the In-School survey in Wave 1. 'Smoke, 'drunk', and 'danger' are dummy variable obtained from the three questions "During the past twelve months, how often did you" separately on 'smoking cigarettes', 'get drunk', and 'do something dangerous because you were dared to' in the In-School survey of AddHealth.

6 Discussion

Previous attempts to examine the effects of peers' parents on students' outcomes find mixed results. For example, Hsieh and Lee (2016) and Hsieh and Lin (2017) estimate spillovers *among* peers. They authors include the education and job status of peers' mothers as control variables and find insignificant effects on students' GPA. In contrast, Patacchini and Zenou (2016) investigate the effects from having same-race friends. They control for education status of either peers' fathers or peers' mothers, depending on who is the interviewee in the survey, and find positive effects. To complement previous works on peers' parents, this paper differentiates the effects of peers' fathers from that of peers' mother, and relaxes the assumption that the spillover is the same for male and female students. When compared to the pooled estimates across gender, I do find that the effects from peers' parents of same-gender are understated, whereas the effects from peers' parents of opposite-gender are overstated.

6.1 Mechanisms

The same-gender and opposite-gender spillover from peers' parents operate through different channels. Parents' choice and direct peer effects matter for the opposite-gender effects. First, the significance of parents' choice in determining a student's outcomes is well-documented. In affecting social networks which their children belong to, parents choose neighborhood, school and peer groups that are best suited to their children (Black, 1999; Bayer et al., 2007; Agostinelli, n.d.). Second, peers affect each other (Manski, 1993b). As shown in Section 4.1, when the proxies of parents' choice and direct peer effects are included, the effects of peers' parents of the opposite gender are explained away.

Taking into accounts these factors, what remain unexplained can be attributed to social influence. Interestingly, only the size of the same-gender spillover remains large and significant. The spread of information about the ideal behaviors of a gender group is a plausible reason.

This resembles the analysis of Akerlof and Kranton (2000) regarding how gender identity shapes behaviors. When social influence is taken into account, an individual conforms to the 'appropriate' standards to avoid losing utility. Indeed, the group 'identity' creates an extra cost in utility maximization in which the cost is generated via the deviation from the 'ideal' behaviors (Akerlof and Kranton, 2002; Ghiglino and Goyal, 2010).

In the current context, the assimilation of the 'ideal behaviors' can go directly through interaction or indirectly via social networks. For direct influence from peers' parents on a child, my result coincides with previous findings on the effect of role models (Bettinger and Long, 2005; Dee, 2007; Griffith, 2014; Bosma et al., 2012). They find that the presence of same gender role models has positive impacts on school performance of a student. One explanation for why interaction only matters for the same-gender pairs is that individuals of the same gender tend to engage in common activities that result in more frequent contacts. This enhances the transmission of influences through changing preferences, sharing information and human capital transfer (Chung, 2000). This also explains the different results for own parents shown in Table 5 because one's own father and mother can affect their children via daily contacts.

Even without direct interaction, the diffusion of beliefs and attitudes can work through social networks. In a field experiment, Avvisati et al. (2013) randomly assign parents to engage in parent-school meetings and learn how to help their children on schoolwork. Not only students from the treated families show significant improvement in attitudes and behaviors (such as truancy and work efforts), students from the untreated families also show improvements if they are the classmates of the treated students. In the current context, the unobserved factors affect male and female students differently. As Chung (2000) suggests, information about the same social group can be more precise about the benefits and costs of an action.

One concern to the above discussion is that contextual factors that are not related to social influences may explain the spillover from peers' parents. For example, more-educated parents who are active in participating in school activities and teacher-parent meetings exert influences on school policies, and thus benefit their child's classmates. Fruehwirth (2016) finds that the reading score of students improves because teachers increase their unpaid preparation time and instruction hours when the parents of their students are more educated. However, contextual factors should not generate the heterogeneity by gender and family background.

7 Conclusion

Social interactions affect our decisions and behaviors through imitation of behaviors, the pressure to conform, or exposure to new information. This paper takes a novel perspective by looking at the diffusion of influence from individuals with indirect social ties. Indeed, the indirect ties in the current discussion, peers' parents, can exert significant influences through inter-generational transmission. By demonstrating the outcomes of certain actions, adults affect young people not only directly through interactions, but also indirectly through transmitting information via social networks.

In this paper, I find that the educational background of peers' parents significantly affects the college attainment of a child. My results show that peer effects and the choice by parents on neighborhood are important in explaining the spillover from peers' parents. These two factors completely explain the effect of peers' parents of the opposite gender. In contrast, peers' parents of the same gender exert influence independent of the direct peer effect. This suggests either greater same gender interactions that allow for the transmission of information, or contributes to the efficacy of role model influence.

Exposure to educated adults in social networks can influence the educational choice of a child. This evidence on the 'exposure effect' echos a recent work by Chetty et al. (2018) who find a strong association of black boy's future income with the presence of black fathers at the census-tract level. Moreover, one way to promote educational achievement of disadvantaged students may have to do with increasing exposure to educated adults. As I show in the paper, this is especially important for children from single parent families who are most vulnerable.

My results also speak to the literature on tracking (Garlick, 2018). To the extent that children for most of their friendships based on who are 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 background may not be a zero-sum game.

References

Agostinelli, Francesco, "Investing in Childrens Skills: An Equilibrium Analysis of Social Interactions and Parental Investments."

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.

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.

Avvisati, Francesco, Marc Gurgand, Nina Guyon, and Eric Maurin, "Getting parents involved: A field experiment in deprived schools," *Review of Economic Studies*, 2013, *81* (1), 57–83.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan, "A unified framework for measuring preferences for schools and neighborhoods," *Journal of Political Economy*, 2007, *115* (4), 588–638.

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.

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

Black, Sandra E, "Do better schools matter? Parental valuation of elementary education," *Quarterly Journal of Economics*, 1999, *114* (2), 577–599.

_ , 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.

Blume, Lawrence E, William A Brock, Steven N Durlauf, and Yannis M Ioannides, "Identification of Social Interactions," 2010.

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

Bosma, Niels, Jolanda Hessels, Veronique Schutjens, Mirjam Van Praag, and Ingrid Verheul, "Entrepreneurship and role models," *Journal of Economic Psychology*, 2012, *33* (2), 410–424.

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, "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, 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.

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 effect of public housing demolition on labor market outcomes of children," 2017.

Coleman, James S, "Social capital in the creation of human capital," *American Journal of Sociology*, 1988, *94*, S95–S120.

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.

Echenique, Federico and Roland G Fryer Jr, "A measure of segregation based on social interactions," *Quarterly Journal of Economics*, 2007, *122* (2), 441–485.

fei Lee, Lung, "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.

Fruehwirth, Jane, "Your Peers Parents: Understanding Classroom Spillovers from Parental Education through Teaching Practices, Parental Involvement and Skill," 2016.

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.

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

Ghiglino, Christian and Sanjeev Goyal, "Keeping up with the neighbors: social interaction in a market economy," *Journal of the European Economic Association*, 2010, *8* (1), 90–119.

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, "The strength of weak ties: A network theory revisited," *Sociological Theory*, 1983, pp. 201–233.

Granovetter, Mark S, "The Strength of Weak Ties1," *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.

Hanushek, Eric A, John F Kain, and Steven G Rivkin, "New evidence about Brown v. Board of Education: The complex effects of school racial composition on achievement," *Journal of Labor Economics*, 2009, *27* (3), 349–383.

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.

— **and Stefano Mosso**, "The economics of human development and social mobility," *Annual Review of Economics*, 2014, *6* (1), 689–733.

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.

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.

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

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.

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.

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.

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.

Olivetti, Claudia, Eleonora Patacchini, and Yves Zenou, "Mothers, Friends and Gender Identity," 2015.

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

Patacchini, Eleonora and Yves Zenou, "The strength of weak ties in crime," *European Economic Review*, 2008, *52* (2), 209–236.

— **and** — , "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.

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.

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

Tümen, Semih, "Career choice and the strength of weak ties," *Central Bank Review*, 2017, *17* (3), 91–97.

Wahba, Jackline and Yves Zenou, "Density, social networks and job search methods: Theory and application to Egypt," *Journal of Development Economics*, 2005, *78* (2), 443–473.

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

Tables and Results

Table 1: 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 IV applies.

Table 2: Positive correlation between college graduation and friends' family background

	Male			Female		
	With college	No college	t-stat.	With college	No college	t-stat.
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: 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.

Table 3: Control variables reduce the size of the effects from peers' parents

	(1)	(2)	(3)	(4)	(5)	(6)
Peers' Father (College)	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' Mother (College)	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 Effect					X	X
Direct Peer Effect##						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

***, ** and * represent 1%, 5% and 10% significance level respectively.

Dependent variable in all regressions is college attainment of students. A network is defined as a school. Spillover from peers' parents is measured by the proportion of peers' father/mother who are college-grad.

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. Direct peer effect is measured by the average characteristics (GPA, race, gender, age and single parent status) of peers, and outdegree (by race).

Table 4: Family background and direct peer effects explain away the opposite-gender spillover

	(1)	(2)	(3)	(4)	(5)	(6)
Male: Peers' Father (College)	1.154***	1.063***	0.826***	0.713***	0.646***	0.460***
	(0.0923)	(0.103)	(0.107)	(0.108)	(0.115)	(0.120)
Male: Peers' Mother (College)	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***
Female: Peers' Father (College)	1.125***	0.976***	0.668***	0.552***	0.395***	0.162
	(0.0857)	(0.0930)	(0.0969)	(0.0984)	(0.106)	(0.111)
Female: Peers' Mother (College)	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 Effect					X	X
Direct Peer Effect##						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

***, ** and * represent 1%, 5% and 10% significance level respectively.

Dependent variable in all regressions is college attainment of students. A network is defined as a school. Spillover from peers' parents is measured by the proportion of peers' father/mother who are college-grad.

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. Direct peer effect is measured by the average characteristics (GPA, race, gender, age and single parent status) of peers, and outdegree (by race).

Table 5: Influence from peers' parents is smaller than peers and own parents

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

***, ** 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 male student, 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 female student, 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 1 standard deviation (0.55) in average peers' GPA.

Table 6: Gender-specific effects remain robust after selection-correction

	Exogenous link	Endogenous link		
		d-dimensional latent factors		
	(1)	(2)	(3)	(4)
		$d = 1$	$d = 2$	$d = 3$
Male: Peers' Father (College)	0.467***	0.470***	0.460***	0.466***
	(0.120)	(0.120)	(0.120)	(0.119)
Male: Peers' Mother (College)	-0.146	-0.147	-0.162	-0.159
	(0.119)	(0.119)	(0.119)	(0.119)
Female: Peers' Father (College)	0.165	0.164	0.151	0.163
	(0.111)	(0.111)	(0.112)	(0.111)
Female: Peers' Mother (College)	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 college attainment of students. All regressions include standard demographic variables, controls for outdegree(by race), as well as grade and network fixed effects. Controls of family background include occupation and education of father and mother, and a single parent indicator; community characteristics include crime rate (county level), median household income (block level) and race dispersion (block level). A network is defined as a school. Spillover from peers' parents is measured by the proportion of peers' father/mother who are college-grad.

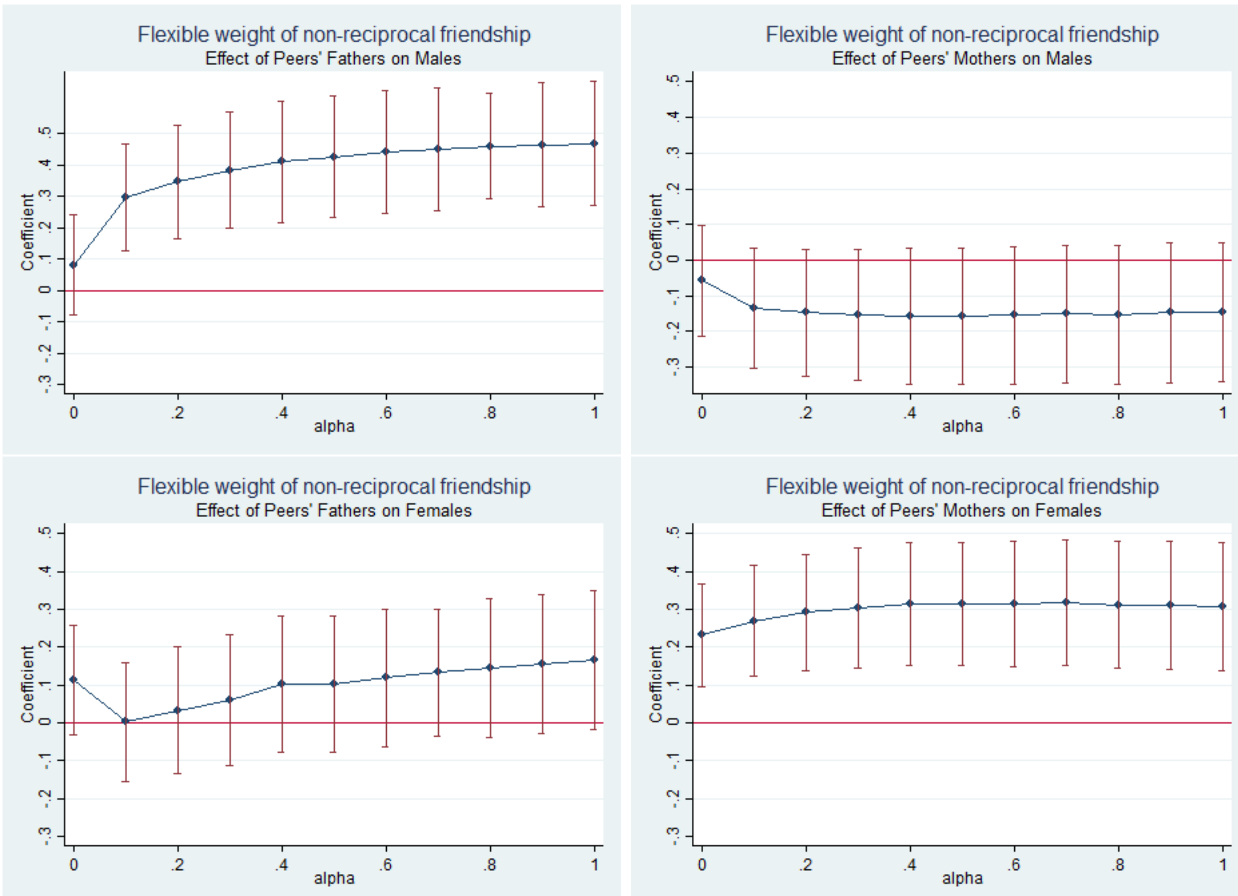

Figure 1: The plots show the changes of the four main coefficients when adjusting the weight (α) on non-reciprocal friendships. Error bar of each dot represents the 90% confidence interval. When α approaches 0, less weights are assigned to non-reciprocal links. When $\alpha = 0$, all links are reciprocal friendships, whereas $\alpha = 1$ resembles the weight used in the main analysis. The estimates are indeed stable across the weighting methods. One exception is the coefficient of the effects of peers' father on males (Top-left panel), when only reciprocal links are counted, the spillover become small and insignificant.

Table 7: Heterogeneity by family background (Linear Probability Model)

	(1)	(2)	(3)	(4)	(5)	(6)
	Two-Parent family					
VARIABLES	Both college	Either college	Neither college	Single FATHER	Single MOTHER	Both absence
Peers' Fathers on Males	0.0813	0.207***	0.167***	-0.00939	0.131*	0.0672
	(0.0668)	(0.0644)	(0.0538)	(0.133)	(0.0703)	(0.128)
Peers' Mothers on Males	-0.0575	-0.0826	0.0126	0.0519	0.0203	-0.112
	(0.0668)	(0.0637)	(0.0529)	(0.129)	(0.0664)	(0.135)
Peers' Fathers on Females	0.0105	0.00529	0.0565	0.0863	0.219***	-0.187
	(0.0608)	(0.0640)	(0.0462)	(0.166)	(0.0645)	(0.173)
Peers' Mothers on Females	0.0953	0.219***	0.0546	0.302*	0.0466	-0.0288
	(0.0594)	(0.0609)	(0.0437)	(0.173)	(0.0554)	(0.141)
Observations	1,237	1,291	2,925	224	1,428	294

Note: 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 with all the control variables included in the main analysis. A linear probability model is employed for the ease of interpretation.

Table 8: 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

The sample size for two-parent families is 5,453. The responses are recorded in the Wave 1 In-Home survey of AddHealth.

The care indexes are obtained using the principal-factor method.

Table 9: 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
Peers' Fathers on Males	0.187***	0.146***	0.189***
	(0.0435)	(0.0385)	(0.0435)
Peers' Mothers on Males	-0.0506	-0.0443	-0.0458
	(0.0409)	(0.0371)	(0.0409)
Peers' Fathers on Females	0.0253	0.0226	0.0282
	(0.0366)	(0.0349)	(0.0367)
Peers' Mothers on Females	0.113***	0.106***	0.114***
	(0.0333)	(0.0324)	(0.0333)
CareIndex	0.0586***	0.00998	0.0546***
	(0.0147)	(0.0108)	(0.0151)
CareIndex*(Peers' Fathers on Males)	-0.0972**	-0.0260	-0.102**
	(0.0482)	(0.0502)	(0.0499)
CareIndex*(Peers' Mothers on Males)	0.0318	0.0450	0.0178
	(0.0462)	(0.0487)	(0.0472)
CareIndex*(Peers' Fathers on Females)	-0.00669	0.0350	-0.0137
	(0.0409)	(0.0470)	(0.0425)
CareIndex*(Peers' Mothers on Females)	-0.0216	-0.0104	-0.0229
	(0.0383)	(0.0473)	(0.0394)
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

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.

Note: A linear probability model is employed. The sample is constrained to two-parent families because responses are skipped for students from single parent families.

Table 10: Correlation matrix of college completion and other outcomes

	college	GPA	smoke	drunk	danger
college	1.000				
GPA	0.409	1.000			
smoke	-0.168	-0.193	1.000		
drunk	-0.085	-0.140	0.468	1.00	
danger	-0.061	-0.057	0.227	0.213	1.00

Data source: 'Smoke, 'drunk', and 'danger' are dummy variable obtained from the three questions "During the past twelve months, how often did you" separately on 'smoking cigarettes', 'get drunk', and 'do something dangerous because you were dared to' in the In-School survey of AddHealth.

The table shows that negative behaviors are not strongly correlated with college attainment. Regressing these other outcomes on the average college attainment of peers' parents can test the hypothesis that behavioral change is one mechanism.

Table 11: Education background of peers' parents does not affect negative behaviors

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	GPA	GPA	smoke	smoke	drunk	drunk	danger	danger
Homogeneous effect								
Peers' Father (College)	0.0970**		-0.0160		0.000718		-0.00781	
	(0.0395)		(0.0262)		(0.0240)		(0.0281)	
Peers' Mother (College)	0.0634*		0.00849		0.00412		-0.0290	
	(0.0358)		(0.0239)		(0.0219)		(0.0255)	
Heterogeneous effect								
Male: Peers' Father (College)		0.152***		-0.0395		-0.0479		-0.0558
		(0.0537)		(0.0354)		(0.0327)		(0.0383)
Male: Peers' Mother (College)		-0.0240		0.0226		0.0142		-0.0416
		(0.0521)		(0.0345)		(0.0318)		(0.0372)
Female: Peers' Father (College)		0.0506		0.0253		0.0537*		0.0416
		(0.0506)		(0.0331)		(0.0306)		(0.0359)
Female: Peers' Mother (College)		0.131***		0.00675		4.65e-05		-0.0181
		(0.0463)		(0.0305)		(0.0281)		(0.0330)
Spillover effect	0.436***	0.436***	0.411***	0.392***	0.439***	0.428***	0.165***	0.162***
	(0.0178)	(0.0178)	(0.0196)	(0.0195)	(0.0185)	(0.0185)	(0.0217)	(0.0217)
Constant	3.500***	3.545***	-0.119	0.226	-0.109	0.0746	0.664**	0.816***
	(0.430)	(0.430)	(0.283)	(0.282)	(0.260)	(0.260)	(0.304)	(0.305)
Observations	7,399	7,399	6,985	6,985	6,985	6,985	6,985	6,985
R-squared	0.279	0.279	0.173	0.193	0.225	0.231	0.105	0.109

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

A simple model without friendship formation is used for all regressions. For dichotomous outcomes from Column (3) to (8), a linear probability model is employed.

A 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:

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

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 $R_g = (A_0^{-1} + l_g'l_g)^{-1}$

B Convergence Diagnosis

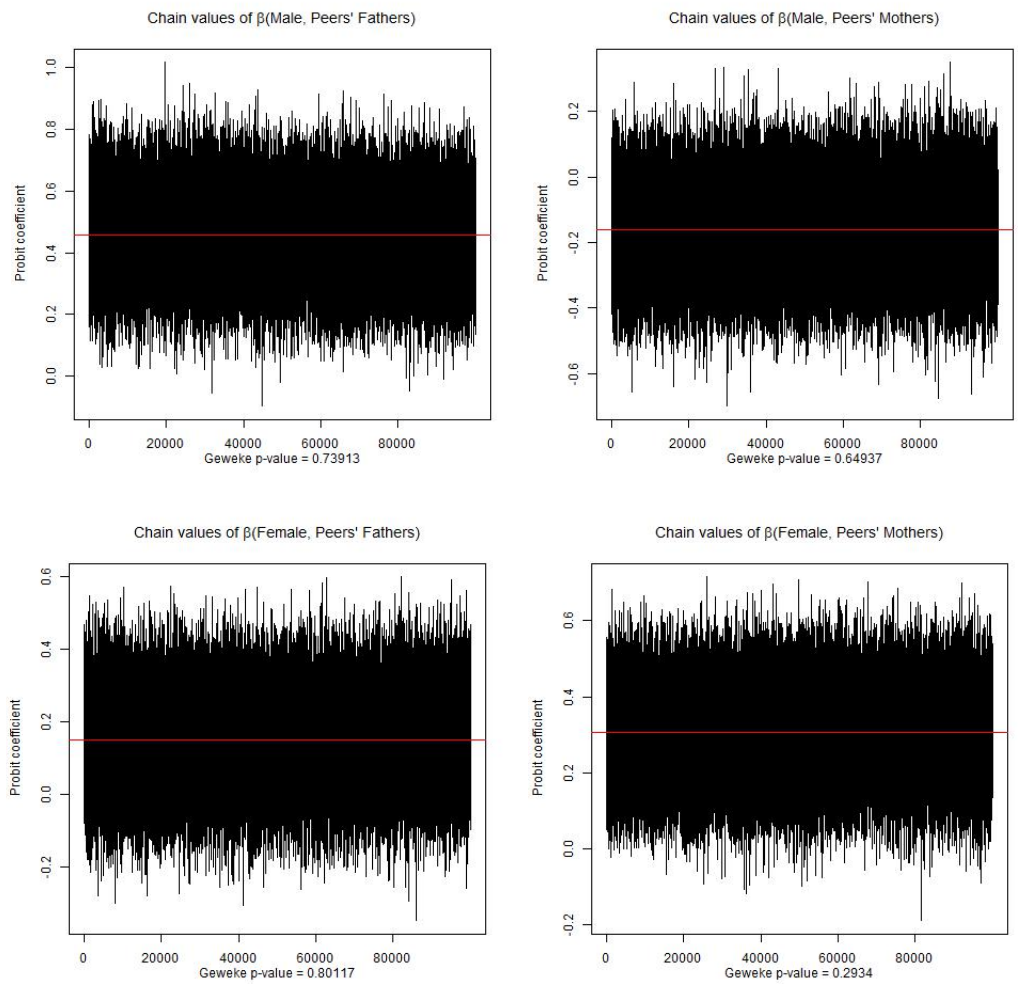

Figure 2: The figures show the chain values of the four variables of interest. Convergence is confirmed by Geweke (1992)'s diagnostic that mimics a simple two-sample test of means between the first 10% and the last 50% of the chain values.

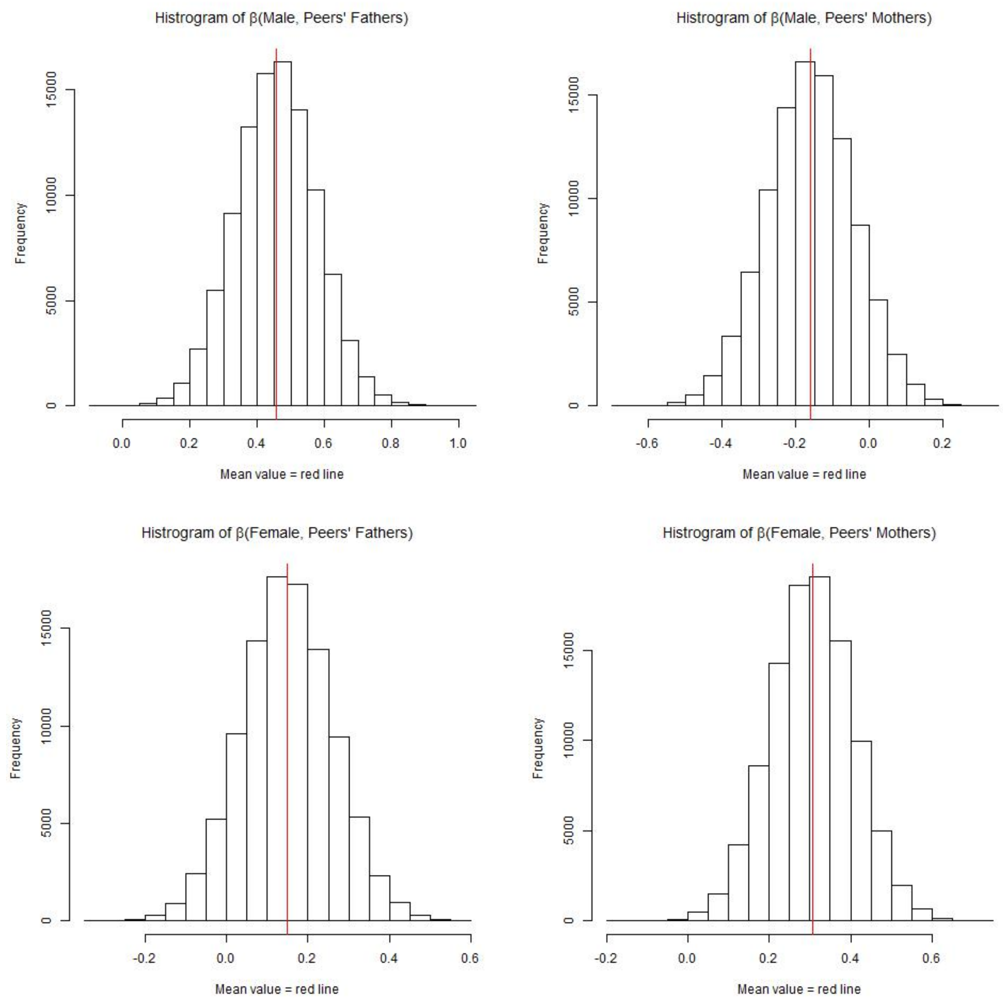

Figure 3: The figures show the histograms of the draws of the four variables.