Economics Bulletin

Volume 34, Issue 3

Network attributes and peer effects

Xu Lin
Wayne State University

Abstract

This paper investigates how network attributes affect the strength of peer influences in adolescents' academic achievement and smoking behaviors. The results indicate that for both GPA and smoking, endogenous peer effects are stronger for network groups with larger size, higher network density or reciprocal link density, while endogenous interaction effects are weaker for network groups with larger non-white or black proportion. And the impact of gender composition on the strength of peer effects is small. Grouping all network groups together may mask the important heterogeneity of peer influences along these dimensions.

I thank an anonymous referee for very helpful comments. 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 to 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.
Citation: Xu Lin, (2014) "Network attributes and peer effects", *Economics Bulletin*, Vol. 34 No. 3 pp. 2060-2079.
Contact: Xu Lin - xulin@wayne.edu.
Submitted: November 22, 2013.   **Published:** September 30, 2014.

1 Introduction

There are numerous studies on peer influences for various behaviors and outcomes, including students' academic achievement (Calvó-Armengol *et al.* 2009; Lin 2010; Sacerdote 2001; Zimmerman 2003); tobacco, alcohol and substance use (Fletcher 2010, 2011; Gaviria and Raphael 2001; Lundborg 2006; Powell *et al.* 2005); school enrollment decisions (Bobonis and Finan 2009; Lalive and Cattaneo 2009); obesity outcomes (Fortin and Yazbeck 2011), and the like. Most of these studies have focused on the average magnitude of peer influences for the whole sample, ignoring the possible heterogeneity of peer influences along various dimensions.

Recently, a few studies have started to explore the possible non-linearity or heterogeneity nature of peer effects along several lines of individual characteristics, such as gender and race (Hanushek *et al.* 2009; Hoxby 2000; Lavy and Schlosser 2011). Most of these studies find that the magnitudes of peer influences vary significantly along these dimensions. For instance, intra-gender peer effects are found to be stronger than cross-gender effects for the participation of retirement plan (Duflo and Saez 2002); the time spent on jobs (Kooreman 2007); the problem of truancy (Soetevent and Kooreman 2007); and tobacco, alcohol and drug use (Clark and Lohéac 2007). While for for students' alcohol expenditure, Kooreman (2007) shows that cross-gender effects are stronger than intra-gender effects. For racial effects, Hoxby (2000) shows that intra-race peer effects are stronger than inter-race. Nakajima (2007) finds that intra-race peer influences are generally positive and significant in students' smoking behaviors while inter-race effects are not.

However, the possible heterogeneity nature of peer effects along the more aggregated dimension — network or group attributes — has seldom been investigated. Arguably, the magnitudes and patterns of social interactions could vary significantly across social groups with different sizes, different interaction densities, and/or different degrees of homogeneity with regard to races, etc. As a matter of fact, several papers, including Boucher *et al.* (2014) and Lee (2007), demonstrate that group size affects the within group social interaction pattern and therefore, variations in group size can provide valuable information for identification of endogenous social interaction effect. However, they do not explicitly estimate the impact of group size on the magnitude of social interactions across network groups. More importantly, they do not consider the fact that if group sizes are too diverse, the peer effect parameter may no longer be homogenous across the whole sample.

To fill this gap in the literature, this study employs the spatial autoregressive (SAR) model in Lee *et al.* (2010) to investigate the relationship between the aggregated network attributes and the strength of social interactions in student academic achievement and smoking behaviors, using the National Longitudinal Study of Adolescent Health (Add Health) data. Instead of estimating the model for the whole sample, I divide the full sample into several subsamples based on various network properties, including network size, network density, network homogeneity in terms of racial composition, and so on, and then estimate the model for each of these subsamples. By comparing the estimated peer effect coefficients across these subsamples, we can see how different group attributes affect the magnitudes of peer influences.

The results indicate that, for both GPA and smoking, endogenous peer effects are stronger for network groups with larger size, higher network density or reciprocal link density, while endogenous interaction effects are weaker for network groups with larger non-white or black proportion. The impact of gender composition on the strength of peer effects is small. Grouping all network groups together, as is often done, may mask the important heterogeneity of peer influences along these dimensions. In addition, these findings point to a potential caveat of the identification strategy proposed in Boucher *et al.* (2014) and Lee (2007) which relies on variations in group size: if the sizes of network groups are too disperse, then it would be inappropriate to treat the endogenous social interaction parameter as homogenous for the whole sample.

2 Model Specification

Following Lin (2010) and Lee *et al.* (2010), the model is specified as:

$$Y_r = \lambda_0 W_r Y_r + X_r \beta_{10} + W_r X_r \beta_{20} + l_r \alpha_r + \epsilon_r, \qquad r = 1, ..., R \qquad (1)$$

where Y_r and X_r are the $m_r \times 1$ vector and $m_r \times k$ matrix of outcomes and characteristics for the m_r members in group r. The $m_r \times m_r$ row-normalized, zero diagonal spatial weights matrix W_r captures the structure of social network, with element $w_{r,ij}$ representing the weight that individual i assigns to peer j. l_r is the m_r-dimensional vector of ones. λ_0 captures the effect of peers' outcomes, i.e. endogenous social effect, β_{20} represents the effect of peers' characteristics, i.e. contextual social effect, whereas α_r represents the group fixed effect, capturing the confounding effect caused by common factors facing the group members. The error terms are assume to be innovations with $Var(\epsilon_r) = \sigma_0^2 I_{m_r}$.

The SAR model differs from the conventional linear-in-means model in the measurements of peer variables. In the linear-in-means model, peer outcomes are measured by group mean outcome, and peer characteristics are measured by group mean characteristics. Both terms are constant across group members and are linearly dependent, thus λ_0 and β_{20} cannot be separately identified, which is the "reflection problem" (Manski 1993).[1] The SAR model breaks down the linear dependency between endogenous effect and contextual effect by introducing individual specific peer measurements. Specifically, peer outcomes are measured by the weighted average of peer (e.g. friend) outcomes, W_rY_r, and peer characteristics are measured by the weighted average of peer (e.g. friend) characteristics, W_rX_r. Hence, if an instrumental variable can be found for the endogenous variable W_rY_r, then the model is identified and the "reflection problem" resolved. As demonstrated in Bramoullé et al. (2009), Lee et al. (2010), and Lin (2010), among others, the incomplete network structure, where not all individuals are friends of each other, provides a set of natural exclusion restrictions for the instrumental variables. For instance, consider an intransitive triad in a network, a, b and c, where persons a and b are friends, and b and c are friends, but a and c are not friends. Then individual $c's$ characteristics can be used to instrument for individual $b's$ outcome that shows up on the right-hand side of person $a's$ equation, since person c directly affects person b (as they are friends) but only affects person a indirectly (as they are not friends). In addition, to separate social interaction effect from non-social interaction effect, i.e. correlated effect, which can cause correlations in group members' outcomes even in the absence of peer effect, I employ the group fixed effect strategy to control for the confounding effect caused by the common factors facing the group members, such as school policy, teacher quality, and the like. The group fixed effect strategy, along with an extensive set of individual characteristics

[1]In particular, the linear-in-means model is given by: $y_{ir} = \lambda_0 E(y_r|r) + \beta_{10}x_{ir} + \beta_{20}E(x_r|r) + \epsilon_{ir}$. From the reduced form: $y_{ir} = \beta_{10}x_{ir} + \frac{\lambda_0\beta_{10}+\beta_{20}}{1-\lambda_0}E(x_r|r) + \epsilon_{ir}$, it can been seen that only some combination of λ_0 and β_{20}, $(\lambda_0\beta_{10} + \beta_{20})/(1-\lambda_0)$, is identified.

and contextual effects in the model, help reduce the unobserved within group heterogeneity to a minimal level.

To eliminate the group fixed effect and estimate the model, Lee *et al.* (2010) consider the de-group-mean transformer matrix

$$J_r = I_{m_r} - \frac{1}{m_r} l_r l_r'$$

where I_{m_r} is identity matrix of dimension m_r. Multiplying this matrix on both sides of Equation (1), we can get

$$\widehat{Y}_r = \lambda_0 J_r W_r \widehat{Y}_r + \widehat{X}_r \beta_{10} + J_r W_r \widehat{X}_r \beta_{20} + \widehat{\epsilon}_r \tag{2}$$

where $\widehat{Y}_r = J_r Y_r, \widehat{X}_r = J_r X_r$ and $\widehat{\epsilon}_r = J_r \epsilon_r$.

Note that the new error terms, $\widehat{\epsilon}_r$, are heteroskedastic, with a singular variance-covariance matrix of rank $(m_r - 1)$:

$$Var(\widehat{\epsilon}_r) = J_r J_r' \sigma_0^2 = J_r \sigma_0^2 \tag{3}$$

To get rid of the dependence among the observations, consider the orthogonal matrix of J_r, $[F_r, H_r]$, where F_r corresponds to the eigenvalues of one and H_r corresponds to the zero eigenvalues. Multiplying Equation (2) by F_r', we get

$$Y_r^* = \lambda_0 W_r^* Y_r^* + X_r^* \beta_{10} + W_r^* X_r^* \beta_{20} + \epsilon_r^* \tag{4}$$

Note that $Y_r^* = F_r' \widehat{Y}_r$, and $\epsilon_r^* = F_r' \widehat{\epsilon}_r$ are $(m_r - 1) \times 1$, $X_r^* = F_r' \widehat{X}_r$ is $(m_r - 1) \times k$, and $W_r^* = F_r' W_r F_r$ is $(m_r - 1) \times (m_r - 1)$. Most importantly, $Var(\epsilon_r^*) = \sigma_0^2 I_{m_r^*}$ and $m_r^* = m_r - 1$.

Denote $\mathcal{X}_r^* = (X_r^*, W_r^* X_r^*)$, and $\beta = (\beta_1', \beta_2')$. The log likelihood function for group r is

$$\begin{aligned} \ln L_r &= -\frac{m_r^*}{2} \ln(2\pi\sigma^2) + \ln |I_{m_r^*} - \lambda W_r^*| \\ &\quad -\frac{1}{2\sigma^2} [(I_{m_r^*} - \lambda W_r^*) Y_r^* - \mathcal{X}_r^* \beta]' [(I_{m_r^*} - \lambda W_r^*) Y_r^* - \mathcal{X}_r^* \beta] \end{aligned} \tag{5}$$

which can be written in terms of the original variables as

$$\begin{aligned} \ln L_r &= -\frac{(m_r - 1)}{2} \ln(2\pi\sigma^2) - \ln(1 - \lambda) + \ln |I_{m_r} - \lambda W_r| \\ &\quad -\frac{1}{2\sigma^2} [(I_{m_r} - \lambda W_r) Y_r - \mathcal{X}_r \beta]' J_r [(I_{m_r} - \lambda W_r) Y_r - \mathcal{X}_r \beta] \end{aligned} \tag{6}$$

where $\mathcal{X}_r = (X_r, W_r X_r)$.

For the whole sample, $\ln \mathcal{L}_n = \sum_{r=1}^{R} \ln L_r$. Maximum likelihood procedure can then be performed.

In this study, groups are defined as school-grade. Peers are specified as friends from the same group and assigned equal weight in the W_r.[2] For instance, if individual i lists person j as one of his/her 6 friends, then the (i, j) element of W_r will be $1/6$. To evaluate the impact of network attributes on the strength of social influences, I divide the whole sample into serval subsamples according the following criteria respectively: network size, network density[3], reciprocal link density[4], network homogeneity such as percentages of non-white, black and male in the network. In particular, for each network attribute, say network size, I consider the cases of two subsamples as well as three subsamples. Specifically, each group is assigned to one of the two subsamples (large and small size subsamples) based on whether or not its group size is in the top 50% of all groups. And for the three subsamples (large, medium and small size subsamples) case, each group is assigned based on whether its group size is in the top 1/3, middle 1/3 or bottom 1/3 among all groups. Model (1) is then estimated for each of these subsamples. By comparing the estimated endogenous effects across these subsamples, we can see how the network attributes affect the magnitude of social interactions.

3 Data Summary

Add Health survey covers students in grades 7-12 from a nationally representative sample of 132 schools during the 1994-95 school year. This study is based on Wave I in-school survey, since it covers everyone who attends the sampled school, meaning an individual's peers are also likely in the sample. The survey covers over 90,000 students, providing information on demographics, family background, as well as various activities including smoking and academic performance. The most unique feature about Add Health is that

[2]In this study, "networks" or "groups" refer to the school-grade, "peers" refer to friends.

[3]Network density is defined as the number of links in the network divided by the total possible number of links, where the total possible number of links is $10m_r$, as the maximum number of friendship nominations in Add Health is restricted to 10.

[4]Reciprocal link density is given by the number of reciprocal links in the network divided by the total possible number of links.

each respondent was asked to identify up to 5 male and 5 female friends, and linkable friend identification numbers are available.[5]

The final sample for GPA consists of 49,559 individuals from 486 network groups. And there are 53,529 individuals and 488 network groups in the final sample of smoking. Summary statistics of the whole sample are reported in Table 1. The mean GPA is 2.872 out of 4. And an average student smoked 3.371 times per month during the past year. On average, the respondents are 15 years old and have stayed in the current school for 2.6 years. 54.7% of the sample are female. For race composition, White, Black, Asian, Hispanic and other race account for 60.8%, 15.9%, 6.2%, 11.6% and 5.5%, respectively. And 55.4% of the sample participate in some sport club. 76.0% of the sample live with both parents. For mother's education, 44.2%, 31.3% and 9.0% of the respondents' mothers have an education level higher than high school, high school and below high school, respectively. For mother's occupation, following Lin (2010), I consider 5 categories: mother on professional job which includes teacher, doctor, lawyer, etc., mother staying home, mother on welfare, mother on other job and missing information, each accounts for 28.1%, 20.0%, 0.6%, 37.2% and 7.7% of the sample, respectively.

Table 2 summarizes the distribution of the network properties, including network size, network density, network homogeneity, and the like. As can be seen, the network groups are quite heterogenous in terms of these attributes. For instance, for the GPA sample, the top 50% subsample have a group size greater than 86.5, the top 1/3 subsample have a group size over 120.5, while the bottom 1/3 subsample have a group size less than 56. And the network density for the top 50% subsample is greater than 0.332, for top 1/3 subsample is over 0.369, and for the bottom 1/3 subsample is less than 0.282. The difference in non-white proportion is also striking, with the cutoff points for the top 50%, top 1/3 and the bottom 1/3 subsamples being 0.271, 0.435 and 0.193, respectively.

To address the possible endogeneity of group attributes, I explore the observed heterogeneity, such as family background and average demographics of the students, across networks with different attributes.[6] Table 3 presents the

[5]The restriction of 5 male and female friends affects only a small fraction of our sample, as less than 10% of the sample listed the maximum of five male or female friends. Therefore, the impact of this restriction on the estimation results should not be a significant concern.

[6]I thank an anonymous referee for pointing this out.

summary statistics for the group mean characteristics for the whole sample, the subsample with smaller group size and the subsample with larger group size. As can been seen, except for two variables, i.e. years in school and sport club membership, all variables have overall similar summary statistics across networks with different group sizes. Furthermore, on average, students from the subsample with larger group size appear to nominate more friends (3.52) than students from the subsample with smaller group size (2.92).[7] These findings suggest that the differences in the strength of peer effect across networks with different network attributes reflect the differences in the social interaction patterns across networks, e.g. denser networks exude more pressure, instead of representing systematic differences across networks with different attributes.

4 Empirical Results

Results based on two sub-samples are reported in Table 4. As can be seen, for both GPA and smoking, the endogenous interaction effects are stronger for the subsample with higher network size, network density, reciprocal link density, or male proportion. In particular, for GPA, the most striking differences are exhibited by the sub-samples with different network density and reciprocal link density, with the estimated peer effect for the top 50% subsample greater than that of the other subsample by 43.3% and 35.9%, respectively, while the difference in the two sub-samples based on gender proportion is small. And the estimated peer influence parameter for the large group subsample is also greater than that of the small groups by 15.4%. For smoking, gender proportion is also the attribute that generates the least difference, while the other attributes all generate significant impacts on the strength of social interaction. In particular, the estimated peer effect for the large group subsample is greater than that of the small groups by 31.1%. These findings point to a potential caveat of the identification strategy proposed in Boucher *et al.* (2014) and Lee (2007): if the sizes of network groups are too disperse, then it would be erroneous to treat the endogenous peer effect parameter as homogenous for the whole sample. In terms of racial composition, for both

[7] I perform the same analysis based on other group attributes, such as network density, reciprocal link density and the like, and find similar patterns. Further, these patterns hold for the three subsamples. These results are available upon request.

GPA and smoking, the sub-samples with smaller proportion of non-white or black show stronger endogenous effects, with the differences in smoking being more striking.

To further check the validity of the results, I run a falsification test using randomly assigned peers.[8] In this specification, I randomly assign 0 or 1 to the elements of the spatial weights matrix, while maintaining the same interaction density as the original network. As shown in Table 5, none of the estimated endogenous effect coefficient is significant. Therefore, the estimation results in the current study do not appear to be driven by unobserved factors.[9]

Table 6 shows the results based on three subsamples. It can be seen that for both GPA and smoking, the strengths of social interactions are monotonically increasing in network size, network density, or reciprocal link density, while monotonically decreasing in non-white or black proportion. In particular, the difference in the peer effects among the subsamples with different group sizes become more substantial: for GPA, the difference between the top 1/3 and bottom 1/3 subsamples is 52.8%, while for smoking, the difference is 49.5%. Again, the difference for the sub-samples based on gender composition is small.

5 Concluding Remarks

This paper investigates the possible heterogeneity of endogenous social effect along the aggregated dimension of network attributes, an under-explored property in the literature. I find that for both GPA and smoking, endogenous peer effects are stronger for network groups with larger size, higher network density or reciprocal link density, while endogenous interaction effects are weaker for network groups with larger non-white or black proportion. The impact of gender composition on the strength of peer effects is small. Grouping all network groups together may mask the important heterogeneity of peer influences along these dimensions. Studies hinge on variations in group size for identification of peer effect need to keep an appropriate balance between group size dispersion and group size homogeneity, as the magnitude

[8] I am grateful to an anonymous referee for suggesting this test.

[9] I also perform a falsification test for the case of three subsamples and find similar patterns. These results are available upon request.

of peer effect parameter in small groups may be very different than that in large groups.

References

Bobonis, G. and F. Finan (2009). Neighborhood peer effects in secondary school enrollment decisions. *The Review of Economics and Statistics*, 91, 695-716.

Boucher, V., Y. Bramoullé, H. Djebbari and B. Fortin (2014), "Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation." Journal of Applied Econometrics, 29(1): 91-109.

Bramoullé, Y., H. Djebbari and B. Fortin (2009), "Identification of peer effects through social networks." Journal of Econometrics 150: 41-55.

Calvó-Armengol, A., E. Patacchini and Y. Zenou (2009), "Peer Effects and Social Networks in Education." Review of Economic Studies, 76: 1239-1267.

Clark, A. and Y. Lohéac (2007). It wasn't me. It was them! Social influence in risky behaviors among young adults, *Journal of Health Economics*, 26, 763-784.

Duflo, E. and E. Saez (2002). Participation and investment decision in a retirement plan: the influence of colleagues' choices. *Journal of Public Economics* 85, 121-148.

Fletcher, J. M. (2010). Social interactions and smoking: Evidence using multiple student coherts, instrumental variables, and school fixed effects. *Health Economics* 19, 466-484.

____,(2011). Peer influences on adolescent alcohol consumption: evidence using instrumental variables/fixed efect approach. *Journal of Population Economics*, 25, 1265-1286.

Fortin, B. and M. Yazbeck (2011), "Peer Effects, Fast Food Consumption and Adolescent Weight Gain." Working paper, Department of Economics, Université Laval.

Gaviria, A., and S. Raphael (2001), "School-Based Peer Effects and Juvenile Behavior." Review of Economics and Statistics, Vol. 83, No. 2: 257-268.

Hanushek, E. A., J. F. Kain, J. M. Markman and S. G. Rivkin (2003), "Does peer ability affect student achievement?" Journal of Applied Econometrics 18: 527-544.

Hanushek, E. A., J. F. Kain and S. G. Rivkin (2009), "New Evidence about Brown v. Board of Education: The Complex Effects of School Racial Composition on Achievement." Journal of Labor Economics, Vol. 27, No.3: 349-383.

Hoxby, C. (2000), "Peer effects in the classroom: Learning from gender and rave variation." NBER Working Paper 7867.

Kooreman, P. (2007). Time, money, peers, and parents: Some data and theories on teenage behavior. *Journal of Population Economics* 20, 9-33.

Lalive, R. and M.A. Cattaneo (2009). Social interactions and schooling decisions. *The Review of Economics and Statistics*, 91, 457-477.

Lavy V. and A. Schlosser (2011), "Mechanisms and Impacts of Gender Peer Effects at School." NBER Working Paper 13292.

Lee L. (2007), "Identification and Estimation of Spatial Econometric Models with Group Interactions, Contextual Factors and Fixed Effects." Journal of Econometrics 140, no. 2: 333-374.

Lee L., X. Liu and X. Lin (2010), "Specification and Estimation of Social Interaction Models with Network Structures." Econometrics Journal, Vol. 13, Issue 2: 145-176.

Lin, X. (2010), "Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables." Journal of Labor Economics 28: 825-860.

Lundborg, P. (2006). Having the wrong friends? Peer effects in adolescent substance use. *Journal of Health Economics* 25, 214-233.

Manski, C. (1993), "Identification of Endogenous Social Effects: The Reflection Problem." Review of Economic Studies 60: 531-542.

Nakajima, R. (2007). Measuring peer effects on youth smoking behavior. *Review of Economic Studies*, 74, 897-935.

Powell, L. M., J. A. Tauras, and H. Ross (2005). The importance of peer effects, cigarette prices and tobacco control policies for youth smoking behavior. *Journal of Health Economics* 24, 950-968.

Sacerdote, B. (2001), "Peer effects with random assignment: Results for dartmouth roommates." The Quarterly Journal of Economics 116(2): 681-704.

Soetevent, A. R. and P. Kooreman (2007). A discrete choice model with social interactions with an application to high school teen behaviors. *Journal of Applied Econometrics* 22, 599-624.

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

Table 1. Sample Summary Statistics

Variable	Mean	Std. Dev.	Min	Max
GPA	2.872	0.788	1	4
Smoking	3.371	8.649	0	30
Age	14.923	1.673	10	19
Years in school	2.551	1.445	1	6
Male	0.453	0.498	0	1
Female	0.547	0.498	0	1
White	0.608	0.488	0	1
Black	0.159	0.366	0	1
Asian	0.062	0.240	0	1
Hispanic	0.116	0.320	0	1
Other race	0.055	0.228	0	1
Sport	0.554	0.497	0	1
Live with both parents	0.760	0.427	0	1
Not live with both parents	0.240	0.427	0	1
Mom education less than HS	0.090	0.287	0	1
Mom education HS	0.313	0.464	0	1
Mom education more than HS	0.442	0.497	0	1
Mom education missing	0.092	0.289	0	1
Mom on professional job	0.281	0.449	0	1
Mom staying home	0.200	0.400	0	1
Mom on other job	0.372	0.483	0	1
Mom on welfare	0.006	0.079	0	1
Mom job missing	0.077	0.266	0	1

Note:
1. The summary statistics shown in table are based on the GPA sample. Those based on the smoking sample are similar.
2. The variables in italics are the omitted categories in estimation.
3. Professional job includes: doctor, lawyer, scientist, teacher, executive, director and the like; Staying home mothers include those who are retired, homemaker or do not work.

Table 2. Distribution of Network Attributes

	bottom 1/3	top 1/2	top 1/3
GPA Sample:			
Network size	56.000	86.500	120.500
Network Intensity	0.282	0.332	0.369
Reciprocal Link Density	0.137	0.160	0.182
Non-white Proportion	0.193	0.271	0.435
Black Proportion	0.015	0.058	0.142
Male Proportion	0.419	0.451	0.480
Smoking Sample:			
Network size	65.000	95.000	125.000
Network Intensity	0.295	0.344	0.386
Reciprocal Link Density	0.145	0.170	0.187
Non-white Proportion	0.188	0.271	0.453
Black Proportion	0.014	0.058	0.142
Male Proportion	0.419	0.450	0.477

Note:
Columns 2, 3 and 4 show the cutoff points for the bottom 1/3, top 1/2 and top 1/3 sub-samples based on various network attributes, respectively.

Economics Bulletin, 2014, Vol. 34 No. 3 pp. 2060-2079

Table 3. Sample Summary Statistics by Group Size

Variable	Whole Sample		Small Sample		Big Sample	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Age	14.868	0.139	14.884	0.163	14.852	0.109
Years in school	3.046	0.628	3.631	0.697	2.461	0.486
Male	0.452	0.083	0.453	0.099	0.450	0.064
Black	0.171	0.172	0.178	0.189	0.163	0.149
Asian	0.052	0.139	0.037	0.142	0.066	0.130
Hispanic	0.104	0.152	0.107	0.171	0.100	0.127
Other race	0.058	0.117	0.060	0.144	0.056	0.080
Sport	0.594	0.108	0.639	0.144	0.550	0.038
Live with both parents	0.758	0.109	0.757	0.144	0.759	0.051
Mom education less than HS	0.092	0.132	0.098	0.165	0.086	0.085
Mom education more than HS	0.429	0.077	0.421	0.102	0.438	0.034
Mom education missing	0.092	0.115	0.094	0.143	0.091	0.073
Mom on professional job	0.274	0.095	0.271	0.126	0.278	0.040
Mom on other job	0.368	0.078	0.360	0.106	0.376	0.026
Mom on welfare	0.007	0.069	0.008	0.079	0.006	0.057
Mom job missing	0.079	0.111	0.080	0.143	0.078	0.063
Number of friends	3.220	0.549	2.920	0.629	3.520	0.318

Note:
1. These summary statistics are for the group mean characteristics based on the GPA sample. Those based on the smoking sample are similar.
2. Professional job includes: doctor, lawyer, scientist, teacher, executive, director and the like; Staying home mothers include those who are retired, homemaker or do not work.

Table 4. Results Based on Two Subsamples

	GPA Sample			Smoking Sample		
	λ_{SG}	λ_{LG}	Diff(%)	λ_{SG}	λ_{LG}	Diff(%)
Network size	0.231	0.266	15.4	0.299	0.393	31.1
	(0.012)	(0.006)		(0.011)	(0.006)	
Network Intensity	0.211	0.302	43.3	0.324	0.417	28.7
	(0.008)	(0.008)		(0.007)	(0.007)	
Reciprocal Link Density	0.220	0.298	35.9	0.304	0.424	39.6
	(0.008)	(0.008)		(0.007)	(0.007)	
Non-white Proportion	0.293	0.226	-22.8	0.418	0.305	-26.9
	(0.008)	(0.007)		(0.007)	(0.007)	
Black Proportion	0.275	0.243	-11.5	0.407	0.337	-17.2
	(0.008)	(0.007)		(0.007)	(0.007)	
Male Proportion	0.255	0.263	3.0	0.354	0.390	10.0
	(0.007)	(0.008)		(0.007)	(0.007)	

Note:
1. λ_{SG} and λ_{LG} are the estimated endogenous effect for the bottom 50% and top 50% sub-samples based on various network attributes, respectively. The difference is calculated by $(\lambda_{LG} - \lambda_{SG})/\lambda_{SG} \times 100\%$.
2. All estimated coefficients are significant at the 1% level. Standard errors are in parentheses.
3. Other controls in the models are own characteristics X as listed in Table 1, contextual effects WX, and group fixed effect.
4. For the whole sample, the estimated endogenous effect for GPA is 0.259, for smoking is 0.374, both significant at the 1% level.

Table 5. Falsification Test Based on Two Subsamples

	λ_{SG}	λ_{LG}
Network size	-0.004	0.009
	(0.015)	(0.008)
Network Intensity	0.011	0.003
	(0.010)	(0.009)
Reciprocal Link Density	-0.009	-0.009
	(0.010)	(0.010)
Non-white Proportion	-0.011	-0.009
	(0.010)	(0.009)
Black Proportion	-0.006	0.004
	(0.010)	(0.009)
Male Proportion	0.011	0.002
	(0.009)	(0.010)

Note:

1. λ_{SG} and λ_{LG} are the estimated endogenous effect for the bottom 50% and top 50% sub-samples based on group size, respectively, for the GPA sample.
2. None of the estimated coefficients is significant. Standard errors are in parentheses.
3. Other controls in the models are own characteristics X as listed in Table 1, contextual effects WX, and group fixed effect.

Table 6. Results Based on Three Subsamples

	λ_{SG}	λ_{MG}	λ_{LG}	Diff(%)
GPA Sample:				
Network size	0.174	0.266	0.266	52.8
	(0.019)	(0.010)	(0.007)	
Network Intensity	0.209	0.245	0.313	49.7
	(0.010)	(0.008)	(0.010)	
Reciprocal Link Density	0.208	0.245	0.325	56.0
	(0.010)	(0.008)	(0.010)	
Non-white Proportion	0.289	0.269	0.219	-24.3
	(0.010)	(0.009)	(0.009)	
Black Proportion	0.295	0.263	0.224	-24.1
	(0.011)	(0.008)	(0.009)	
Male Proportion	0.239	0.272	0.261	9.3
	(0.009)	(0.009)	(0.010)	
Smoking Sample:				
Network size	0.259	0.375	0.388	49.5
	(0.017)	(0.009)	(0.006)	
Network Intensity	0.302	0.382	0.424	40.2
	(0.009)	(0.008)	(0.009)	
Reciprocal Link Density	0.281	0.367	0.443	57.7
	(0.009)	(0.008)	(0.010)	
Non-white Proportion	0.436	0.370	0.278	-36.2
	(0.010)	(0.008)	(0.008)	
Black Proportion	0.423	0.380	0.313	-26.0
	(0.010)	(0.008)	(0.009)	
Male Proportion	0.349	0.381	0.386	10.6
	(0.009)	(0.008)	(0.009)	

Note:

1. λ_{SG}, λ_{MG} and λ_{LG} are the estimated endogenous effect for the bottom 1/3, middle 1/3 and top 1/3 sub-samples based on various network attributes, respectively. The difference is calculated by $(\lambda_{LG} - \lambda_{SG})/\lambda_{SG} \times 100\%$.
2. All estimated coefficients are significant at the 1% level. Standard errors are in parentheses.
3. Other controls in the models are own characteristics X as listed in Table 1, contextual effects WX, and group fixed effect.

4. For the whole sample, the estimated endogenous effect for GPA is 0.259, for smoking is 0.374, both significant at the 1% level.