Alma Mater Studiorum Università di Bologna
Archivio istituzionale della ricerca

Residential choices of young Americans

(Article begins on next page)

Residential Choices of Young Americans

Eleonora Patacchini
Cornell University,
EIEF, CEPR and IZA

Tiziano Arduini
University of Bologna

ABSTRACT

Residential Choices of Young Americans[*]

Using detailed data on a cohort of young Americans who were in their late twenties and early thirties in 2008, we investigate the importance of forces different from economic incentives in nest-leaving decisions. We apply recent methods from social network econometrics to identify the importance of peers net of confounding factors. For the entire sample, our findings reveal no evidence of peer effects. Indicators of parenting and the social structure of families appear to be the major factors in the decisions to coreside with parents. However, for those who moved back home after a few years of living alone, we find strong peer effects. These findings are consistent with theories of social influences in peer groups in which peers play a critical role for individuals with time-inconsistent preferences.

Corresponding author:

Eleonora Patacchini
Department of Economics
Cornell University
Ithaca, NY 14853
USA
E-mail: ep454@cornell.edu

Since 2007, the share of young adults aged 18-29 living with their parents has been growing steadily in the United States.[1] Although the dynamics differed by gender and race, the increasing trend was a common factor.

Understanding the reasons why young adults remain at their parents' home is of primary policy concern, since the living arrangements of young adults are closely related to fertility, mobility, and labor market outcomes, and hence are related to economic growth. The rising number of young Americans living with their parents in recent years has been attributed to the lower employment prospects and lower wages in the years surrounding the Great Recession.[2,3]

The marked heterogeneity of young adults' decisions within gender, race, household income and marital status categories, however, suggests that other forces such as differences in attitudes in family environments and peer pressure may be at work.[4] Although peer effects have been shown to be important determinants of behavior in a variety of contexts, the housing market is a notable exception.[5] The existing studies on the importance of social interactions in this area of research are extremely limited (see Ioannides, 2012 for a critical survey).[6]

This paper contributes to this literature. It does so by providing estimates based on novel data and obtained using the most recent econometrics techniques that control for network

---

[1]According to the U.S Census Bureau, between 2007 and 2011, the number of young adults living at home rose from 4.7 million to 5.9 million.

[2]See, e.g. Dyrda, Kaplan, and Rios-Rull (2012) and the references therein. Kaplan (2012) builds a structural model and shows that moving back to the parental home acts as insurance against labor market shocks.

[3]Even before the start of the latest recession, employment prospects and associated wages were on the decline for young adults in North America, especially for men.

[4]There is a long-standing economic literature on the importance of demographic and economic factors for residential choices, which is particularly florid for Southern European countries where youths remain at their parents home longer that their counterparts in Scandinavian Europe, the United Kingdom and the United States. See Kiernan (1986) for an international comparison of young adults' living arrangements in Denmark, Great Britain, and the United States; Yi et al. (1994) for a comparison of age-specific net rates of leaving home for men and women in China, Japan, South Korea, the United States, Sweden and France; and Iacovou (2002) for living arrangements of young adults in Europe and the United States. See Manacorda and Moretti (2006), Giuliano (2007), and Chiuri and Del Boca (2010) for the possible consequences of late emancipation of young adults in Southern Europe on their labor market outcomes and on fertility rates.

[†]Examples include education, crime, labor market, fertility, obesity, productivity, participation in welfare programs, risky behavior (for surveys, see Glaeser and Scheinkman, 2001; Moffitt, 2001; Durlauf, 2004; Ioannides and Loury, 2004; Jackson, 2009; Ioannides, 2012).

[δ]A recent contribution is Patacchini and Zenou (2015).

endogeneity. In fact, the most challenging issue faced by all studies using social network data to identify peer effects is that individuals sort into groups in a non-random way. If the variables that drive this process of selection are not fully observable by the researcher, then potential correlations between (unobserved) group-specific factors and the target regressors are a major sources of bias. To address this issue, most of the existing papers (see, in particular, Bramoullé et al., 2009; Calvó-Amengol et al., 2009; Lin, 2010; Lee et al., 2010) use the architecture of the networks by introducing netmork xed effects in the econometric equation. The underlying assumption is that the unobservable factors that drive friendship formation are common to all individuals belonging to the same network. This means that it is assumed that the structure of interactions is (conditionally) exogenous. However, if there are individual-level unobservables that drive both network formation and outcome choices, this strategy will not work.[7] Because of a failure to account for similarities in unobserved characteristics, similar behaviors might mistakenly be attributed to peer influence when they are simply due to similar unobserved characteristics.

In this paper, we explicitly model network formation and estimate a model of link formation and outcomes using a Bayesian approach.[8] By doing so, we account for the possible presence of unobservable individual characteristics affecting both network formation and outcome decisions. The importance of this methodological innovation is confirmed by the fact that the results are dramatically different when we account for network formation.

We use data from the U.S. National Longitudinal Survey of Adolescent Health (AddHealth). This data contains unique information on parents and friends during adolescence for a cohort of young adults who were in their late twenties and early thirties in 2008. This cohort has been followed through the transition into young adulthood with four in-home interviews. The most recent was in 2008, when respondents were 24-34 years old. We use Wave I data (i.e. when individuals were aged 11-21) to obtain a detailed picture of the family and social environments during adolescence. Since the median age of leaving the parental home is around 21-22 for females and 22-24 for males (see, e.g., Iacovou, 2002), we then use the follow-up data in

---

[7]For a general discussion and overview on these issues, see Blume et al. (2011), Goldsmith-Pinkham and Imbens (2013), Graham (2015), and Jackson et al. (2015).

[8]A similar modeling approach is used by Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2014).

2002-2003 (i.e. at Wave III when individuals were aged 18-28) to derive information on nest-leaving decisions. In our sample, about 14,000 students are coresidents with parents in Wave I and about half of them leave the nest in Wave III (excluding homeless and those with missing values). Using the information at Wave IV, we can also identify a small sample of non-coresident individuals who moved back home. This sample consists of slightly fewer than 600 individuals. Particularly important for our study is that the richness of the AddHealth information provides us with a set of "nonstandard" variables to account for the heterogeneity of our sample in terms of parenting and the social structure of the families.

Once we control for unobserved factors driving friendship choices, our findings reveal no effect of peers' behavior on individual behavior for the entire sample. Outside of economic incentives, own family experiences (most notably the quality of parenting and the social structure of families) are the major driving factors. When we restrict our attention to individuals who moved back home, our analysis reveals strong peer effects. These findings are consistent with the view that the peer influence is crucial in shaping behavior for people with problems of self-control and time-inconsistent preferences (see, e.g. Battaglini, Benabou and Tirole, 2005). Nest-leaving behavior does not seem to be an exception.

Adamopoulou and Kaya (2013) find evidence of peer effects in nest-leaving decisions using the same data source (AddHealth). However, they extract different information from the dataset[9] and do not account for endogeneity of friendship formation. In addition, they do not consider the sub-sample of boomerang kids.

The paper unfolds as follows. In the next section, we describe our data and empirical strategy. In Section 3, we present our empirical results and robustness checks. In Section 4, we conclude.

1   Data

Our data source is the National Longitudinal Survey of Adolescent Health (AddHealth), which is a nationally representative survey of more than 90,000 adolescents that began with in-school

---

[9] See footnote 14.

questionnaires administered to U.S. adolescents in grades 7-12 in 1994-1995.[10] The in−school survey contains questions on respondents' demographic and behavioral characteristics, education, family background and friendship. Importantly for the purpose of this paper, this survey also contains unique information on friendship relationships. The friendship information is based upon actual friends' nominations. Pupils were asked to identify their best friends from a school roster (up to five males and five females).[11] The uniqueness of this information lies in the fact that, by matching the identification numbers of the friendship nominations to respondents' identification numbers, one can obtain information on the characteristics of nominated friends.[12]A subsample of these adolescents (around 20,000) were also asked to complete in-home interviews and were followed in three subsequent waves. The in−home survey contains questions relating to more sensitive individual and household information. The household roster at Wave I allows us to identify the other coresident members of the households and subsequent questions in the follow-up waves allows us to identify precisely who moved out and back in through ages 24-32. At Wave I, we define an individual as a coresident if at least one of his/her household members is identified as either father, mother's husband, mother's partner, mother, father's wife, or father's partner. Otherwise, we define an individual as a non-coresident. At Waves III and IV we use the direct question: "Where do you live now? That is, where do you stay most often?", with possible answers: parents' home, another person's home, your own place (apartment, house, trailer, etc.), group quarters (dormitory, barracks, group home, hospital, communal home, prison or penitentiary, etc.), or homeless (you have no regular place to stay).[13]

---

[10]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.

[11]The limit in the number of nominations is not binding, not even by gender. Less than 1 percent of the students in our sample list ten best friends, less than 3 percent list five males and roughly 4 percent list five females.

[12]The other existing survey data collecting information on social contacts (e.g. NSHAP, BHPS, GSOEP) are "ego networks". They contain a list of the contacts each respondent declares with few demographic characteristics (gender, relationship with respondent, education) of each contact, which are self-reported by the respondent. No extensive interview with each nominated contact is performed.

[13]We exclude those who are homeless or live in group quarters, and also those refusing to answer the question.

If respondents claim to live in their parents[3] home, we call them coresidents. Otherwise, we call them non-coresidents. Using the corresponding information for nominated friends, we are able to calculate, the percentage of coresidents (at Wave III) among each individual's peers (at Wave I).[14]

Our sample consists of respondents who met the following criteria: completed Wave I, Wave III, and Wave IV in-home surveys; lived with at least one parent in Wave I; and listed valid information for at least one friend in Wave I (i.e., the friend has coresidence information and can be tracked in the school roster).[15] Our final sample of students consists of slightly fewer than 3500 individuals, out of which roughly 2000 are non-coresidents and roughly 1500 are coresidents at Wave III.

Given that friendship is a reciprocal relationship, we define i and (as friends if at least one of them name the other as best friend in the nomination list.[16] Figure 1 shows the distribution of students by their number of friends, distinguishing between coresident and non-coresident kids at Wave III. While on average AddHealth students have about 2.5 friends, there is a large dispersion around this mean value.[17] Figure 1 reveals that the distribution is bimodal, with the large majority of students having between one and three friends, and a sizeable fraction with many friends (between nine and eleven). However, the distributions for coresidents and non-coresidents are remarkably similar. A formal comparison of the two distributions does not reject the null hypothesis that the two samples are two random drawings from the same population (the Wilcoxon signed-rank test p-value is equal to 0.4356; the paired samples t-test for equality in means p-value is equal to 0.7769).

[Insert Figure 1 here]

---

[14]Wave III also contains a calendar of geographical mobility listing all previous states of residence and the month and year of each move. Unfortunately, information on coresidents in each location is not reported. Adamopoulou and Kaya (2013) use this data to investigate nest-leaving decisions and assume that the last move to the current address corresponds to individuals moving out alone.

[1†]About 40% of respondents in the AddHealth survey do not list any friend.

[16]We use alternative friend definitions in Section 4.

[17]Note that, when an individual i identifies a best friend (who does not belong to the surveyed schools, the database does not include (in the network of i; it provides no information about (. However, in the large majority of cases (more than 94%), students tend to nominate best friends who are students in the same school and thus are systematically included in the network.

Figure 2 shows the distribution of networks by network size. We again distinguish between coresident and non- coresident kids. One can see that the social circles in our sample are quite small, since the large majority of social networks (more than 75%) have fewer than 15 members. Again, there is a marked similarity between the coresident and non-coresident distributions (the Wilcoxon signed-rank test p-value is equal to 0.3435; the paired samples t-test for equality in means p-value is equal to 0.6667). As a result, this evidence reveals that children who decide to leave the parental home are not different in the number of social contacts from those who do not leave.

[Insert Figure 2 here]

In Table 1, we investigate the presence of other differences in observable characteristics. Table 1 contains a description of the variables used in our study, as well as descriptive statistics on our sample. We display statistics for coresident and non-coresident individuals in different columns. Coresidents are more likely to be male, non-white, and unemployed, and are more likely to have low school grades. They are more likely to come from relatively poor and less-educated families. Interestingly, they also differ from non-coresidents in terms of the social structure of families and parenting. Non-coresidents are more likely to come from families with two parents, from families where parents are married, and from smaller households. They are also likely to have spent more evening time with their parents during adolescence.

[Insert Table 1 here]

1.1 Empirizal Model and Estimation strategy

Let us consider a population of $N = (1, \ldots, n\}$ individuals distributed among K networks. Let n_k be the number of individuals in the kth network, so that $N = \sum_{k=1}^{K} n_k$. Let us denote by $G_k = [g_{i(,k}]$ the adjacency matrix of a network k. It captures the direct connections in this network. Here, two agents i and (are directly connected (i.e. best friends) in k if and only if $g_{i(,k} = 1$, with $g_{i(,k} = 0$ otherwise. We also set $g_{ii,k} = 0$. The set of individual i's best friends (direct connections) is: $N_i(k) = ((/= i \mid g_{i(,k} = 1\}$, which is of size $g_{i,k}$ (i.e. $g_{i,k} = \sum_{(=1}^{n} g_{i(,k}$ is

7

the number of direct links of individual i). In particular, this means that, if **i** and (are best friends, $N_i(k) \neq N_((k)$ unless the graph/network is complete (i.e. each individual is friend with everybody in the network). This also implies that groups of friends may overlap if individuals have common best friends. To summarize, the reference group of each individual i is $N_i(k)$, the set of his/her best friends, which does not include him/herself.

The decision of individual i in network k to leave the parental home, $y_{i,n}$, can be modeled as:

$$y_{i,n} = \$ \frac{1}{g_{i,k}} \sum_{(=1}^{n_n} g_{i(,k} \, y_{(,k} + \sum_{m=1}^{M} \beta_1^m x_{i,n}^m + \frac{1}{g_{i,k}} \sum_{m=1}^{M} \sum_{(=1}^{n_n} 8_m g_{i(,k} x_{(,n}^m + y_k + o_{i,k} \qquad (1)$$

where $\frac{1}{g_{i,h}} \sum_{(=1}^{n_n} g_{i(,k} \, y_{(,k}$ denotes the share of friends that left parental home in individual i's reference group and $\sum_{m=1}^{M} \beta_1^m x_{i,n}^m + \frac{1}{g_{i,h}} \sum_{m=1}^{M} \sum_{(=1}^{n_n} 8_m g_{i(,k} x_{(,n}^m$ reflects the ex ante idiosyncratic heterogeneity of each individual i in terms of one's own characteristics and friends'characteristics, as captured by the set $x_{i,n}^m$ (for $m = 1, \ldots, M$). Finally, y_k captures network-specific unobserved factors (constant over individuals in the same network), which might be correlated with the regressors, and $o_{i,k}$ is a white noise error.[18]

Conformity preferences, which state that the individual wants to minimize the social dis‐tance between herself and her reference group, or learning mechanisms, in which peers are a channel for information sharing, provide the behavioral foundations for this model.[19]

In model (1), \$ represents the endogenous effects, or the impact of one's friends' activities on her choice/outcome for the same activity. In addition, 8 represents the contextual effect, or the extent to which an agent's choice/outcome may depend on the exogenous characteristics of her friends. The vector of network fixed effects y_k captures the correlated effect, whereby agents in the same network may behave similarly because of similar unobserved individual characteristics or a similar environment.

---

[18]In the spatial econometrics literature, model (1) is the so-called spatsal lag model or msued‐regresssve spatsal autoregresssve model (Anselin, 1988) with the addition of a network-specific component of the error term. A maximum likelihood approach is used to estimate $\emptyset$, $\hat{\theta}$, and ϕ jointly (see, e.g. Anselin, 1988).

[19]See Clark and Oswald (1998), Akerlof (1997) and Patacchini and Zenou (2012) for conformism and peer effects and Banerjee (1992), Battaglini et al. (2005) and Moretti (2011) for learning models with peer effects.

1.2 Fdentifization and estimation

A number of papers have dealt with the identification and estimation of peer effects with network data (see, e.g. Bramoullè et al., 2009; Lee, 2007; Liu and Lee, 2010; Lee et al., 2010; Calvó-Armengol et al., 2009). Below, we review the crucial issues and explain how we address them.

Reflection problem In linear-in-means models, simultaneity in the behavior of interacting agents introduces a perfect collinearity between the expected mean outcome of the group and its mean characteristics. Therefore, it is difficult to differentiate between the effect of peers' choice of effort (endogenous effects) and peers' characteristics (contextual effects) that have an impact on their effort choice. Manski (1993) terms this the so-called reflection problem. The reflection problem arises because, in the standard approach, individuals interact in groups - individuals are affected by all individuals belonging to their group and by nobody outside of their group. However, in the case of social networks this is almost never true since the reference group is individual-specific. For example, take individuals i and I such that $g_{il,k} = 1$. Then, individual i is directly influenced by $\bar{y}_{i,k} = \sum_{(=1}^{n_h} g_{i(,k} y_($ while individual l is directly influenced by $\bar{y}_{l,k} = \sum_{l(=1} g_{l(,k} y_($, and there is little chance for these two values to be the same unless the network is complete (i.e. everybody is friend with everybody else).[20]

Correlated effects/Sorting While a network approach allows us to distinguish endogenous effects from correlated effects, it does not necessarily enable us to estimate the causal effect of peers' influences on individual behavior. In most cases, individuals sort into groups non-randomly. For example, students whose parents have lower-than-average educational attainment might be more likely to sort into groups with lower human capital. If the variables that drive this process of selection are not fully observable, potential correlations between (unobserved) group-specific factors and the target regressors are a major sources of bias. The richness of social network data (where we observe individuals over networks) provides a possible solution through the use of netmork xed effects. Network fixed effects are a remedy for the selection bias that originates from the possible sorting of individuals with similar un-

---

[20]Formally, social effects are identified (i.e. there is no reflection problem) if I, G, G^2 and G^3 are linearly independent. The intuition is that the sntransstsvsty sn socsal connectsons in social networks data provide exclusion restrictions to identify endogenous and contextual effects (see, e.g. Bramoullè et al., 2009).

observed characteristics into a network. The underlying assumption is that such unobserved characteristics are common to the individuals within each network.

However, if there are individual-level unobservables that drive both network formation and outcome choice, this strategy fails. For example, one can envision the existence of unobservable (or unmeasurable) factors, such as self confidence or risk aversion, which are possibly relevant both in social contexts and for nest-leaving decisions. Recently, Goldsmith-Pinkham and Imbens (2013)and Hsieh and Lee (2014) highlight the fact that endogeneity of this sort can be modeled. Individual-level correlated unobservables would motivate the use of parametric modeling assumptions and Bayesian inferential methods to integrate network formation with the study of behavior over the formed networks. In this paper, we develop this approach for our context. It is detailed below.

Endogenous network formation As mentioned in the introduction, the most challenging issue faced by all studies attempting to identify peer effects is a possible endogeneity of the network. If there are individual-level unobservables that drive both network formation and outcome choices, the estimates of peer effects are biased.

Let $_{i,k}$ denote an unobserved characteristic of individual i belonging to network k that influences the link formation process. Let us also assume that $_{i,k}$ is correlated with $_{i,k}$ in Model (1) according to a bivariate normal distribution

$$
(_{i,k}, _{i,h}) \sim N @@ \begin{pmatrix} 0 \\ 0 \end{pmatrix} , , \begin{pmatrix} \sigma^2_{s_{i,h}} & \sigma_{s_{i,h} \, i,h} \\ \sigma_{s_{i,h} \, i,h} & \sigma^2_{i,h} \end{pmatrix} , , .
$$

Joint normality implies that the error term $_{i,h}$ in equation (1) can be replaced with its expected value $\sigma_{s_{i,h} \, i,h \, i,h}$, yielding:

$$
y_{i,k} = \$ \frac{1}{g_{i,k}} \sum_{(=1}^{n_h} g_{i(,k} \, y_{(,k} + \sum_{m=1}^{M} \beta^m_1 x^m_{i,k} + \frac{1}{g_{i,k}} \sum_{m=1}^{M} \sum_{(=1}^{n_h} 8_m g_{i(,k} x^m_{(,k} + y_k + \sigma_{s_{i,h} \, i,h} \, _{i,k} + u_{i,k} \quad (2)
$$

where $u_{i,k}$ is now an i.i.d. error term uncorrelated with the $x^m_{i,k}$ and the unobservable $_{i,k}$. Observe that $\sigma_{s_{i,h} \, i,h} \, _{i,k} /= 0$ implies that the network G in model (1) is endogenous.[21]

---

[21]For simplicity, we consider only one unobserved characteristic governing the link formation process. The

Let us thus consider a network formation model based on homophily behaviors, where the variables that explain friendship ties between students i and ι belonging to network k (i.e. $g_{i\iota,k}$) are the distances between them in terms of observed and unobserved characteristics. Let us assume that the probability of two individuals being friends $g_{i\iota,k}$ follows a logit specification of the form

$$P(g_{i\iota,k} = 1|x_{il,k}, x_{\iota l,k}, \eta_{il,k}, \eta_{\iota l,k}) = \Omega\left(\delta_0 + \sum_m |x_{i,k}^m - x_{\iota,k}^m|\delta^m + \delta_2|\eta_{i,k} - \eta_{\iota,k}|\right) \qquad (3)$$

where $\Omega(\cdot)$ is the logistic distribution and δ_0, δ_1, δ_2 are parameters governing friendship formation. Among the observable individual characteristics (x variables), we also include a dummy variable taking a value of 1 if students i and ι reside in the same neighborhood and zero otherwise. Equation (3) explains the link formation process between individuals i and ι in network k by their difference in observable characteristics (i.e. $|x_{i,k}^m - x_{\iota,k}^m|$) and unobservable characteristics (i.e. $|\eta_{i,k} - \eta_{\iota,k}|$). This is a standard model of homophily (see e.g. Currarini et al., 2009, 2010).

Equations (1) and (3) form a structural model of link formation and outcomes. The main advantage of this approach is that possible friendship selection bias on network interactions can be corrected as the network formation is explicitly modeled.

We use this model to study peer effects in nest-leaving decisions and estimate it using the Bayesian method. A Bayesian approach will produce marginal posterior distributions of the parameters, conditioned on the data and the set of individual-level nuisance parameters $\eta_{i,k}$. We can interpret $\eta_{\iota,k}$ as an individual fixed-effect that affects the outcome and also explains the probability of two individuals i and ι being friends. Details on the Bayesian estimation procedure can be found in Appendix B.

2 EmpiriCal results

We present the estimation results of model (1) on the entire sample using a wide range of specifications and various estimation strategies. They are reported in Table 2. The last row of this

introduction of different unobservables simply adds more notation.

table shows the percentage of the variance that is explained by peer effects. We begin in column (1) by showing the raw correlation between the individual probability of leaving the parental home and the share of peers that left home. The correlation is quite high (about 36% in terms of explained variance). When we control for standard individual characteristics (column (2)), the portion of the variance explained falls to 26.5%. In Column (3), we introduce controls for family characteristics, including parenting activities (which are typically unobserved) and indicators for the social structure of families. Interestingly, it appears that those variables have an explanatory power roughly as large as the other economic and demographic factors. The portion of variance attributed to peer effects drops to roughly 17%. However, a correlation between individual and peers' behavior may be due to similar individual and peer characteristics, rather than to peer effects (i.e. endogenous effects). The uniqueness of our data where both respondents and friends are interviewed allows us to control for peers' characteristics, thus disentangling the effects of endogenous from exogenous effects. Column (4) shows that about half of effect attributed to peers' behavior is in fact due to peers' characteristics -the portion of the variance explained falls from 17% to 9%. A remaining concern relates to the presence of unobserved factors. There are two types of unobservables: (i) unobservables that are common to all individuals in a (broadly defined) social circle and (ii) unobservables that are individual-specific. The bi-dimensional nature of network data (we observe individuals over networks) allows us to control for the presence of unobserved factors of type (i) by including network fixed effects.[22] By doing so, we purge our estimates from the effects of unobserved factors that are common among directly and indirectly related individuals. Column (5) reports the results when network fixed effects are included in the model. The percentage of explained variance falls by about 6%, thus revealing the presence of important unobserved factors in each individual's social circle. The presence of type (ii) unobservables is more difficult to address. The application of an econometric strategy able to deal with this issue in our context is an important contribution of this paper. Such a strategy, which is detailed in Section 2.2, consists of simultaneously estimating the outcome equation (2) and the link formation equation (3).

---

[22]This is a pseudo panel data within-group strategy, where the group mean (here network mean) is removed from each individual observation. Network fixed effects therefore are not estimated. They are treated as nuisance parameters.

By explicitly modeling network formation, these estimates correct for possible friendship selection bias. When this method is used, column (6) reveals that the percentage of the variance explained drops to 0.

[Insert Table 2 here]

We report the complete list of estimation results in Table 3. In column 3, it can be seen that the effect of the behavior of peers on nest-leaving decisions is insignificant. In fact, column (2) reveals that the estimated correlation between unobservables in the outcome and link formation equations ($\sigma_{@}$) is different from zero. This suggests that the effects of friends are mainly due to the unobservable individual characteristics that also drive the friendship formation. Therefore, there are no endogenous peer effects. In other words, we observe a correlation between individual decisions to coreside with parents and the share of peers coresiding with parents not because the individual decisions are affected by the decisions of friends but because adolescents in friendship circles share some common unobservable characteristics that make them friends and also drive nest-leaving behavior. (For example, parental attitudes, parental working time, cultural norms, and living standards affect both friendship formation and the choices of living arrangements). The estimation results on our control variables are in line with the descriptive statistics in Table 1. It appears that a higher probability of leaving the parental home is associated with coming from a relatively wealthy and highly-educated family, a family with two parents, a family with married parents, and a family in which parents spend evening time with their children. Interestingly, in terms of magnitude, the impact of indicators of the social structure of families and parenting is non-negligible. It can be compared with the effect of being employed. Indeed, if individuals are employed the probability of leaving parental home is roughly 20% higher. This probability is about 18% higher if parents spend one more evening per week with their children. It is about 23% higher for individuals coming from families with two parents.

[Insert Table 3 here]

13

2.1 Boomerang kids

According to the information collected in the Add Health survey in Wave IV, about 25% of the respondents who declared they were non-coresidents at Wave III moved back in with their parents. What are the reasons that these adolescents, the so called boomerang kids, leave their parents' home and then return a few years later? A conventional explanation would attribute those movements to labor market shocks and to changes in marital status. Finding or losing a job, marriage and marriage termination are the obvious candidates. Table 4 compares characteristics of the young adults that remain non-coresidents at Wave IV with the boomerang kids. The boomerang kids are more likely to be male and to come from wealthier families. However, contrary to expectations, it does not appear that boomerang kids have higher unemployment rate, nor higher separation or divorce rates.

[Insert Table 4 here]

We then investigate whether the boomerang kids are different from the others in terms of social contacts. Figures 3 and 4 show the graphs that correspond to Figures 1 and 2, respectively. We distinguish between coresident and non-coresident kids. Both graphs reveal marked similarities, and formal statistical tests do not detect any difference in the distributions. Therefore, the evidence so far does not help us to understand the differences in behavior between the two samples.

[Insert Figures 3 and 4 here]

Table 5 details the results from repeating our analysis for the boomerang kids. In Table 5, the dependent variable is equal to one if the individual left the parental home in Wave III and returned in Wave IV, and zero otherwise. In other words, we look at peer effects in nest-leaving decisions if an individual is a boomerang kid. The magnitudes and the signs of the coefficients of the control variables are similar to those in Table 3. Table 5 shows significant and strong peer effects. This evidence can be interpreted in light of theories of social influences in peer groups where peers play a critical role for individuals with time-

inconsistent preferences and/or individuals subject to episodes of temptation such as drinking, smoking, drug use, sexual activity, procrastination of effort, etc. In particular, Battaglini, Benabou and Tirole (2005) develop a model of self control in peer groups in which externalities arise endogenously from inferences among peers who observe each other's behavior. If being a boomerang kid signals that an individual has limited willpower, then our results are consistent with this theory. Our data seems to suggest that this seems to be the case. The AddHealth questionnaire contain questions that are commonly used to measure self-control or willpower (see Nagin and Pogarsky, 2001; Fletcher et al., 2009; Wolfe and Hoffmann, 2016; Battaglini et al., 2015).[23] We then regress these alternative measures on a dummy variable taking value one if the individual is a boomerang kid and zero otherwise, controlling for individual and family background characteristics. The results are shown in Table 6. It appears that boomerang kids have significantly lower self-control than the other kids, irrespective of the measure of self-control used.

[Insert Tables 5 and 6 here]

3   Robustness zhezks

In this section, we examine the sensitivity of our results to possible measurement error in the definition of the peer group.

3.1   Undirezted vs direzted networks

Our empirical investigation assumes, so far, that friendship relationships are symmetric, i.e. $g_{i(} = g_{(i}$. Our data, however, make it possible to know exactly who nominates whom in a

---

[23]The precise (Wave I) questions are: "When making decisions, you usually go with your "gut feeling" without thinking too much about the consequences of each alternative?", coded 1= strongly disagree to 5= strongly agree; If you wanted to use birth control, how sure are you that you could stop yourself and use birth control once you were highly aroused or turned on?, coded 1=very unsure to 5=very sure; When you have a problem to solve, one of the first things you do is get as many facts about the problem as possible, coded 1= strongly disagree to 5= strongly agree; When making decisions, you generally use a systematic method for judging and comparing alternatives, coded 1= strongly disagree to 5= strongly agree; How often was the following true during the past week? You had trouble keeping your mind on what you were doing, coded 1= most or all of the time to 4= never or rarely.

network, and we find that 12 percent of relationships in our dataset are not reciprocal. Instead of constructing undirected networks, we will now focus on the analysis of directed networks.

In a directed graph, a link has two distinct ends: a head (the end with an arrow) and a tail. Each end is counted separately. The sum of head endpoints count toward the indegree and the sum of tail endpoints count toward the outdegree. Formally, we denote a link from i to ℓ as $g_{i\ell} = 1$ if ℓ has nominated i as her friend, and $g_{i\ell} = 0$, otherwise. The indegree of student i, denoted by g_i^+, is the number of nominations student i receives from other students, that is $g_i^+ = \sum_\ell g_{i\ell}$. The outdegree of student i, denoted by g_i^-, is the number of friends student i nominates, that is $g_i^- = \sum_\ell g_{\ell i}$. We can thus construct two types of directed networks, one based on indegrees and the other based on outdegrees.

We report in Table 7, panel (a) the results of the estimation of model (1) and (3) when we use these alternative definitions of network links.[24] In column (1) we use the entire sample. In column (2) we restrict the sample to the boomerang kids. Our results are only minimally affected when using alternative network structures. Indeed, we still find that peer influences act as a social multiplier only for those kids that leave the parental home and return soon after.

3.2 Strong ties vs weak ties

Because we observe friendship ties a few years before residing decisions are made, we investigate if the relevant peers are only friends that persist over time. While the AddHealth does not allow us to follow friendship evolution into adulthood, the survey collects friendship nominations in two waves (Wave I and Wave II). This feature of the data allows us to identify friendships that persisted over (at least) one year. We define friends as strong ties if they have nominated each other in both waves; we define friends as weak ties if they have nominated each other in one wave only.[25] Panel (b) of Table 7 reports the results of the estimation of model (1) and (3) when distinguishing between different peer types. The results again show no evidence of peer effects for the entire sample (column 1). When looking at the sub-sample of boomerang kids (column 2), we find that both strong and weak ties are relevant, although the influence of

---

[24]We report the results on the target variables. The complete list of estimation results is available upon request.

[25]A similar peer classification has been used by Patacchini et al. (2016) to study heterogeneous peer effects in education.

weak ties is less than half the effect of strong ties.

[Insert Table 7 here]

4   Conzluding remarks

This paper investigates whether and to what extent factors different from conventional economic variables play a role in explaining differences in living arrangements among younger Americans. Our data provide two unique pieces of information: (i) information on non-standard characteristics of the family environment, and (ii) information on nominated peers. Our econometric methodology allows us to tackle possible endogeneity of peer choice. We find that indicators of parenting and of the social structure of families, which are typically unobserved, explain a non-negligible portion of the variance in young adults' nest-leaving decisions. For the entire sample, we do not find evidence of a social multiplier in nest-leaving decisions. However, for individuals who leave their parents' house but then return, we do find evidence of peer effects. This segment of the young population is not negligible and has been growing over time.[26] Our analysis has some limitations. In our sample, there are relatively few individuals who left their parents' home and then returned. Additionally, time spent with children and the legal status of parents can proxy for a broad set of environmental and behavioral factors. Given these limitations, our findings should be interpreted as a suggestive evidence that factors different from economic incentives, which are usually unobserved, might play an important role in shaping young adults' residential choices. Further investigation, together with additional data, is clearly required to tease out precise policy recommendations.

---

[26]According to the U.S. Census Bureau, the number of young men aged 25-34 who are living with their parents grew by more than 30% between 2005 and 2011 (data source: http://www.census.gov/population/www/socdemo/hh-fam/cps2011.html).

Referenzes

[1] Adamopoulou, E. and Kaya, E. (2013), Young Adults Living with their Parents and the Influence of Peers, Universidad Carlos III de Madrid, Working Paper 13-10.

[2] Akerlof, G. A. (1997), "Social distance and social decisions", Econometrica 65, 1005-1027.

[3] Anselin, L. (1988), Spatial Econometrics, Methods and Models, Dordrech: Kluwer Academics.

[4] Banerjee, A. V. (1992), "A simple model of herd behavior", fhe Quarterly Journal of Economics, 797-817.

[5] Battaglini, M., Benabou, R., and Tirole, J. (2005), "Self-control in peer groups"', Journal of Economic fheory, 123(2), 105-134.

[6] Battaglini, M., Diaz, C. and Patacchini E. (2015), "Self-control in peer groups", manuscript Cornell University.

[7] Blume, L., Brock, W., Durlauf, S., and Ioannides, Y. (2011), "Identification of Social Interactions", Handbook of Social Economics, 1B, 853-964.

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

[9] Calvó-Armengol, A., Patacchini, E., and Zenou, Y. (2009), "Peer effects and social networks in education", Reviem of Economic Studies 76, 1239-1267.

[10] Chib, S. and Greenberg, E. (1996). "Markov chain monte carlo simulation methods in econometrics", Econometric theory, 12(03), 409-431.

[11] Chiuri, M., and Del Boca, D. (2010). "Home-leaving Decisions of Daughters and Sons", Reviem of Economics of the Household, 8, 393-408.

[12] Clark, A. E., and Oswald, A. J. (1998), "Comparison-concave utility and following behaviour in social and economic settings", Journal of Public Economics 70, 133-155.

[13] Currarini, S., Jackson, M.O., and Pin, P. (2009), "An economic model of friendship: Homophily, minorities, and segregation", Econometrica 77, 1003-1045.

[14] Currarini, S., Jackson, M.O., and Pin, P. (2010), "Identifying the roles of race-based choice and chance in high school friendship network formation", Proceedings of the National Academy of Sciences of the USA 107, 4857-4861.

[15] Durlauf, S.E. (2004), "Neighborhood effects", In: Handbook of Regional and Urban Economics Vol. 4, J.V. Henderson and J-F. Thisse (Eds.), Amsterdam: Elsevier Science, 2173-2242.

[16] Dyrda, S., Kaplan, G., and Rios-Rull, V. (2012), "Business Cycles and Household Formation: The Micro vs the Macro Labor Elasticity", NBER Working Paper.

[17] Fletcher, J. M., Deb, Partha, and Sindelar, J. L. (2009), "Tobacco Use, Taxation and self-control in Adolescence," NBER Working Paper 15130.

[18] Glaeser, E.L., and Scheinkman, J. (2001), "Measuring social interactions ", In: Social Dynamics, S.N. Durlauf and H.P. Young (Eds.), Cambridge: MIT Press, 83-132.

[19] Giuliano, P. (2007), "Living Arrangements in Western Europe: Does Cultural Origin Matter?", Journal of the European Economic Association, 5, 927-952.

[20] Goldsmith-Pinkham, P. and Imbens, G. W. (2013), "Social networks and the identification of peer effects", Journal of Business and Economic Statistics 31, 253-264.

[21] Graham, B. S. (2015), "Methods of identification in social networks", National Bureau of Economic Research.

[22] Hsieh, C.-S. and Lee, L. F. (2014), "A social interactions model with endogenous friendship formation and selectivity", Journal of Applied Econometrics.

[23] Iacovou, M. (2002), "Regional Differences in the Transition to Adulthood", Annals of the American Association of Political and Social Science, 580, 40-69.

[24] Ioannides, Y.M. (2012), From Neighborhoods to Nations: The Economics of Social Interactions. Princeton University Press

[25] Ioannides Y.M., and Loury, L. D. (2004), "Job Information Networks, Neighborhood Effects, and Inequality", Journal of Economic Giterature, 42(4), 1056-1093

[26] Jackson, M. O. (2009), "Networks and economic behavior," Annu. Rev. Econ., 1(1), 489-511.

[27] Jackson, M. O., Rogers, B. W., and Zenou, Y. (2015). "The Economic Consequences of Social Network Structure", Available at SSRN.

[28] Kaplan, G. (2012), "Moving Back Home: Insurance Against Labor Market Risk", Journal of Political Economy, 120(3): 446-512.

[29] Kiernan, K. (1986) "Leaving home: living arrangements of young people in six West-European countries," European Journal of Population, Vol. 2, 177-184.

[30] Lee, L-F. (2007), "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects", Journal of Econometrics 140, 333-374.

[31] Lee, L-F., Liu, X., and Lin, X. (2010), "Specification and estimation of social interaction models with network structures", Econometrics Journal 13, 145-176.

[32] Lin, X. (2010), "Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables", Journal of Gabor Economics 28, 825-860.

[33] Liu, X. and Lee, L-F. (2010), "GMM estimation of social interaction models with centrality", Journal of Econometrics 159, 99-115.

[34] Manacorda, M., and Moretti, E. (2006). "Why Do Most Italian Youths Live With Their Parents? Intergenerational Transfers and Household Structure", Journal of the European Economic Association, 4, 800-829.

[35] Manski, C.F. (1993), "Identification of endogenous effects: The reflection problem", Reviem of Economic Studies 60, 531-542.

[36] Moffitt, R. (2001), "Policy interventions low-level equilibria, and social interactions", In: S. Durlauf and P. Young (Eds.), Social Dynamics, Cambridge, MA: MIT Press, 45-82.

[37] Moretti, E. (2011),"Social learning and peer effects in consumption: Evidence from movie sales", fhe Reviem of Economic Studies, 78(1), 356-393.

[38] Nagin, D. S. and Pogarsky, G. (2001), "Integrating celerity, impulsivity, and extralegal sanction threats into a model of generally deterrence: Theory and evidence," Griminology, 39, 865-92.

[39] Patacchini, E., Rainone, E. and Zenou, Y. (2016), "Heterogeneous peer effects in education", Bank of Italy working paper n. 1048.

[40] Patacchini, E., and Zenou, Y. (2015), "Racial identity and education in social networks," Social Netmorks, forthcoming.

[41] Patacchini, E., and Zenou, Y. (2012), "Juvenile delinquency and conformism", Journal of Gam, Economic, and Organization 28, 1-31.

[42] Tierney, L. (1994), "Markov chains for exploring posterior distributions", the Annals of Statistics, 1701-1728.

[43] Yi, Z., Coale, A., Kim Choe, M., Zhiwu, L., and Li, L. (1994), "Leaving the Parental Home: Census-Based Estimates for China, Japan, South Korea, United States, France, and Sweden", Population Studies, Vol. 48, 65-80.

[44] Wolfe E. and J. P. Hoffmann (2016), "On the Measurement of Low Self-Control in Add Health and NLSY79", Psychology Grime and Gam , forthcoming.

Appendi< B: Bayesian Estimation

Prior and Posteriors Distributions

In order to draw random values from the marginal posterior distributions of parameters in Models (2) and (3), we need to set prior distributions of those parameters. Once priors and likelihoods are specified, we can derive marginal posterior distributions of parameters and draw values from them. Given the link formation Model (3), the probability of observing a network k, G_k, is

$$P(G_k|x_{i,k}, x_{(,k}, \ _{i,k}, \ _{(,k}, 6^m, 6_0, 6_2) \ = \ \prod_{i \ (} P(g_{i(,k}|x_{i,k}, x_{(,k}, \ _{i,k}, \ _{(,k}, 6^m, 6_0, 6_2),$$

Let $\beta^\times = (\beta^J_1, 8^J)$, following HL11 our prior distributions are

$$z_{i,r} \ \sim \ N(0, 1)$$
$$c \ \sim \ N_{M+2}(c_0, \Omega_0)$$
$$\$ \ \sim \ U[-n, n]$$
$$\beta^\times \ \sim \ N_{2M+1}(\beta_0, B_0)$$
$$(\varphi^2, \sigma_o) \ \sim \ T N_2(\sigma_0, C_0)$$
$$y_k|\sigma_y \ \sim \ N(0, \sigma_y)$$
$$\sigma_y \ \sim \ IG(\frac{c_0}{2}, \frac{0}{2})$$

where $c = (6^J, 6_0, 6_2)$, $n = \frac{1}{n} - |\$|$, $n = 1/\max(\min(\max_i(\sum_{(} g_{i()}), \max_{(}(\sum_i g_{i()})))$ from Gershgorin Theorem, $U[\cdot]$, $T N_2(\cdot)$ and $IG(\cdot)$ are respectively the uniform, bivariate truncated normal, and inverse gamma distributions. Those distributions depend on hyper-parameters

22

(like β_0) that are set by the econometrician. It follows that the marginal posteriors are

$$P(\xi_k|Y_k, G_k, g) \propto \prod_{k=1}^{K}\prod_{i}^{n_h} \$(\xi_{i,k})P(Y_k, G_k|\xi_k, g)$$

$$P(c|Y_k, G_k) \propto \phi^{M+2}(c, c_0, \Omega_0)\prod_{k=1}^{K} P(G_k|\xi_k, c)$$

$$P(\$|Y_k, G_k, \xi_k, \beta, \sigma_0^2, \sigma_o) \propto \prod_{k=1}^{K} P(Y_k^2|G, \xi_k, \beta^{-1}, \sigma, \sigma_o)_{k\,k}$$

$$P(\beta^\times|Y_k, G_k, \xi_k, \sigma_0^2, \sigma_o, \$) \propto \phi^{2M+1}(\tilde{\beta}, \tilde{B})$$

$$P(\sigma_0^2, \sigma_o|Y_k, G_k, \xi_k, \$) \propto \phi^2((\sigma_0^2, \sigma_o), \sigma_0, C_0)\prod_{k=1}^{K} P(Y_k|G_k, \xi_k, \beta^\times, \sigma_\sigma^2, \sigma_o, \sigma_y)$$

$$P(y_k|Y_k, G_k, \xi_k, \$, \sigma^2, \sigma_o, \sigma_y) \propto \$(y_k, \tilde{y}_{k}, \tilde{M}_k)$$

$$P(\sigma_y|Y_k, G_k, \xi_k, \$, \sigma_o^2, \sigma_o) \propto L\varsigma\left(\frac{\varsigma_0 + r}{2}, \frac{\sigma + \sum_{k=1}^{K} y_k^2}{2}\right)$$

where $g = (c, \$, \beta^\times, \sigma_0^2, \sigma_o, \sigma_y, y)$, $\phi^l(\cdot)$ is the multivariate $l-$ dimensional normal density function, $\phi_T^l(\cdot)$ is the truncated counterpart, $L\varsigma(\cdot)$ is the inverse gamma density function. $\tilde{\beta} = \tilde{B}(B_0^{-1}\beta_0 + \sum_{k=1}^{K} X_k' V_k(S_k Y_k - \sigma_o\xi_k))$, $\tilde{B} = (B_0^{-1} + \sum_{k=1}^{K} X_k' V_k X_k)^{-1}$, $\tilde{y}_k = (\sigma_\sigma^2 - \sigma_o^2)^{-1}\tilde{M}_k I_n^{\iota\,h}(S_{kr}Y_k - \sigma_o\xi_k - X_k^\times\beta^\times)$, and $\tilde{M}_k = (\sigma_y^{-2} + (\sigma^2 - \sigma^2{}^0)^{-1}I_n^{\iota\,h} I_n^{h})^{-1}$, where $V_k = (\sigma^2{}^0 - \sigma_\sigma^2)I_{nh} + \sigma_y^2 I_{nh} I_{nh}'$, where $X_k^\times = (X_k, G_k^\times X_k)$. The posteriors of $\beta^\times, (y_k)$ and σ_y are available in closed forms and a usual Gibbs Sampler is used to draw from them. The other parameters are drawn using the Metropolis-Hastings (M-H) algorithm (Metropolis-within-Gibbs).[27]

Sampling Algorithm

We start our algorithm by picking $(c^{(1)}, \$^{(1)}, \beta^{\times(1)}, \sigma_o^{2(1)}, \sigma_o^{(1)}, \sigma_y^{(1)}, y^{(1)})$ as starting values. For $\beta^{\times(1)}, y^{(1)}, \$^{(1)}$ we use OLS estimates, while we set the variances-covariances $\sigma_o^{2(1)}, \sigma_o^{(1)}, \sigma_y^{(1)}$ at 0.[28] We ought to draw samples of $\xi_{i,k}^t$ from $P(xi_{i,k}|Y_k, G_k, g)$, $i = 1, \cdots, n$. To do this, we first draw a candidate $\tilde{\xi}_{i,k}^t$ from a normal distribution with mean $\xi_{i,k}^{t-1}$, then we rely on a M-H

---

[27]See Tierney (1994) and Chib and Greenberg (1996) for details regarding the resulting Markov chain given by the combination of those two methods.

[28]The algorithm is robust to different starting values. However, speed of convergence may increase significantly.

decision rule: if $\tilde{\imath}^{t}_{i,k}$ is accepted we set $\imath^{t}_{i,k} = \tilde{\imath}^{t}_{i,k}$, otherwise we set $\imath^{t}_{i,k} = \imath^{t-1}_{i,k}$. Once all $\imath_{i,k}$ are sampled, we move to the sampling of β^x. By specifying a normal prior and a normal likelihood we can now easily sample β^t from a multivariate normal distribution. A diffuse prior for σ^2_s allows us to sample it from an inverse chi-squared distribution. We follow the Bayesian spatial econometric literature by sampling \$ from uniform distributions with support $[-n, n]$ as defined above. A M-H step is then performed over a normal likelihood: if accepted, then $\$^{t} = \tilde{\$}^{t}$. For network fixed effects we deal again with normal prior and normal likelihood, so y is easily sampled from a multivariate normal. We sample σ^2, σ_o from a truncated bivariate normal over an admissible region Ξ such that the variance-covariance matrix is positive definite. Acceptance or rejection is determined by the usual M-H decision rule. A detailed step-by-step description of the algorithm is provided below.

Step 1: Sample $\imath^{t}_{k}$ from $P(\imath_{t k}|Y_k, G_k, g)$.

Propose $\tilde{\imath}_k$ drawing each $\tilde{x}i_{i,k}$ from $N(\imath^{(t-1)}_{i,k}, \Xi)$, then set $\imath^{t}_{i,k} = \tilde{\imath}^{t}_{i,k}$ with probability or $\imath^{t}_{i,k} = \imath^{t-1}_{i,k}$ with probability $1 -$ where

$$= \min\left[\frac{P(Y_k|G_k, \tilde{\imath}^{t}_{k}, g^{t-1})}{P(Y_k|G_k, \imath^{t-1}_{k}, g^{t-1})} \cdot \prod_i \frac{P(g \mid \tilde{\imath}^{t}_{i(k}, \imath^{t-1}_{i,k}, (\zeta_k \quad \$(\tilde{\imath}^{t}_{ij k})}{P(g \mid \imath^{t-1}_{i,k}, c) \$(\quad)^{t-1}}\right]^{n_h}$$

Step 2: Sample c^t from $P(c|Y_k, G_k)$.

Propose $\tilde{c}^t$ from $N^{M+2}(c^{t-1}, {}_c\Omega_0)$, then set $c^t = \tilde{c}^t$ with probability $_c$ or $c^t = c^{t-1}$ with probability $1 - {}_c$ where

$$_c = \min\left(\prod_{r=1}^{Y} \frac{P(G_k|Z^{t}_k, \tilde{c}^t)}{P(G_k|\imath^{t}_k, c^{t-1})} \cdot \frac{\phi^{2M+1}(\tilde{c}^t, c_0, \Omega_0)}{\phi^{2M+1}(c^{t-1}, c_0, \Omega_0)}\right)$$

Step 3: Sample $\tilde{\$}^{t}$ from $P(\$|Y_k, G_k, \imath_k, \beta^x, \sigma^2_o, \sigma_o)$.

Propose $\tilde{\$}^{t}$ from $N(\$^{t-1}, {}_\$)$, then set $\$^t = \tilde{\$}^{t}$ with probability $_\$$ or $\$^t = \$^{t-1}$ where

$$_\$ = \min\left[\prod_{h=1}^{K} \frac{P(Y^h|G^h, \imath^{t-}_{h}, \tilde{\$}, \beta^{xt-1}, \sigma^{2t-1}_{..}, \sigma^{t-1}_{..}, \sigma^{t-1}_{y})}{P(Y^h|G^h, \imath^{t-1}_{h}, \$^{t-1}, \beta^{xt-1}, \sigma^{2t-1}, \sigma^{t-1}, \sigma^{t-1})} \cdot I(\tilde{\$} \quad \$ [-n, n])\right]^{t}$$

Step 4: Sample σ^2_o and σ^t_o from $P(\sigma^2, \sigma_o|Y_k, G_k, \$)$.

Propose σ^t_o and σ^t_o from $N^2((\sigma^{2t-1}_0, \sigma^{t-1}_0), {}_\sigma, C_0)$, then set $\sigma^t_0 = \sigma^t_0$ and $\sigma^t_0 = \sigma^t_0$ with

24

probability α_σ or $\sigma_0^t = \sigma_0^{t-1}$ and $\sigma_0^t = \sigma_0^{t-1}$ with probability $1 - \alpha_\sigma$ where

$$\alpha_\sigma = \min\left(\Psi \frac{P(Y_h|G_h, \beta_h^{t-1}, \$^{t-1}, \beta^{\times t-1}, \tilde\sigma_u^t, \tilde\sigma_u^t, \sigma_y^{t-1})}{\prod_{h=1} P(Y_h|G_h, \beta_h^{t-1}, \$^{t-1}, \beta^{\times t-1}, \sigma_u^{t-1}, \sigma_u^{t-1}, \sigma_y^{t-1})} \frac{\Phi_T^2((\sigma_u^t, \sigma_u^t), \sigma_0, Z_0)}{\Phi^2(\sigma_u^{t-1}, \sigma_u^{t-1}, \sigma_0, Z_0)} I((\sigma^t, \sigma^t) \$ \Xi) \right)$$

where Ξ is a region in which the variance-covariance matrix is definite properly.

Step 5: Sample $\beta^{\times t-1}$, y^t and σ_y^t from conditional posterior distributions.

Step 6: Repeat previous steps updating values indexed with t.

In each of the M-H steps (1-4) the algorithm accepts the new random values (proposals) if the likelihood is higher than the current one. In the algorithm, γ, γ_c, γ_σ, and $\gamma_\$$ are tuning parameters chosen by the econometrician. This choice determines the rejection rate of proposals in the M-H steps (1-4). We set a dynamic algorithm for calibrating those tuning parameters so that they converge to the optimal ones. Optimality means that the proposals are accepted about 50% of the time.[29] In our application, convergence is achieved around an acceptance rate of 50% for all of the parameters.[30]

---

[29]The intuition is that if a tuning parameter is too high, the draws are less likely to be within "high density regions" of the posterior and then rejection is too frequent. The "step" is too long and the chain "does not move enough". On the other hand if the "step" is too short, the proposal is more likely to be accepted and the chain "moves too much". Given that we want a mixing chain with a balanced proportion of rejections and acceptances, an optimal step must be chosen. Setting it manually requires a huge amount of time and many manual operations. The dynamic setting of tuning parameters is as follows:

if $t_A/t \le 0.4$ then $\gamma_{t+1} = \gamma_t/1.1$,

if $t_A/t \,c\, 0.6$ then $\gamma_{t+1} = \gamma_t \times 1.1$,

if $0.4 \,)\, t_A/t \,)\, 0.6$ then $\gamma_{t+1} = \gamma_t$,

where t_A is the acceptance rate at iteration t. The procedure decreases the tuning parameter (the "step") when proposals are rejected too frequently, while it increases the tuning parameter when proposals are accepted too frequently. This mechanism guarantees a bounded acceptance rate and convergence to optimal tuning.

[30]Detailed results on the convergences are available upon request.

Figure 1. Distribution of individuals by number of social contacts and residential choices
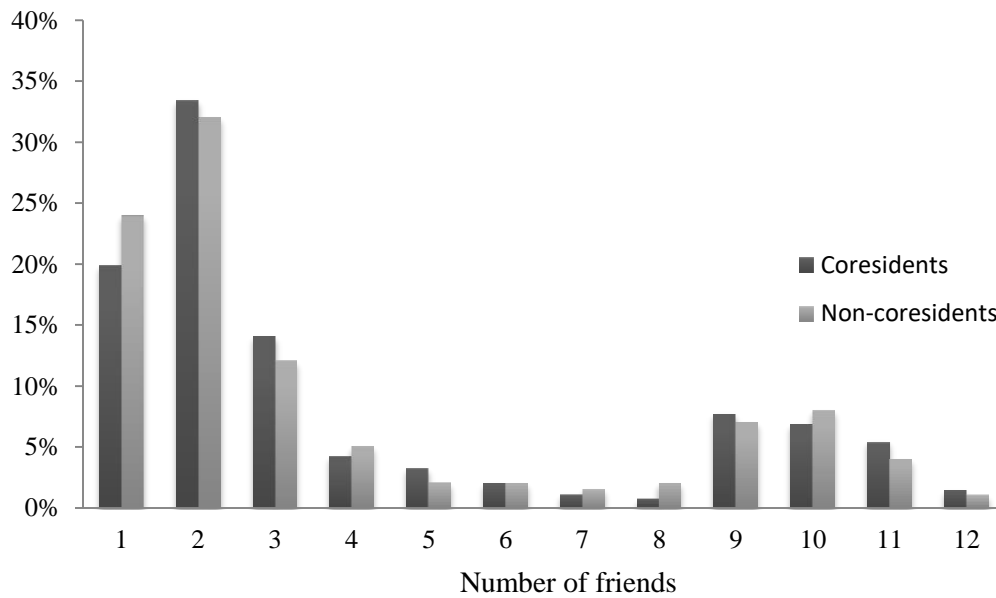

Figure 2. Distribution of networks by network size and residential choices
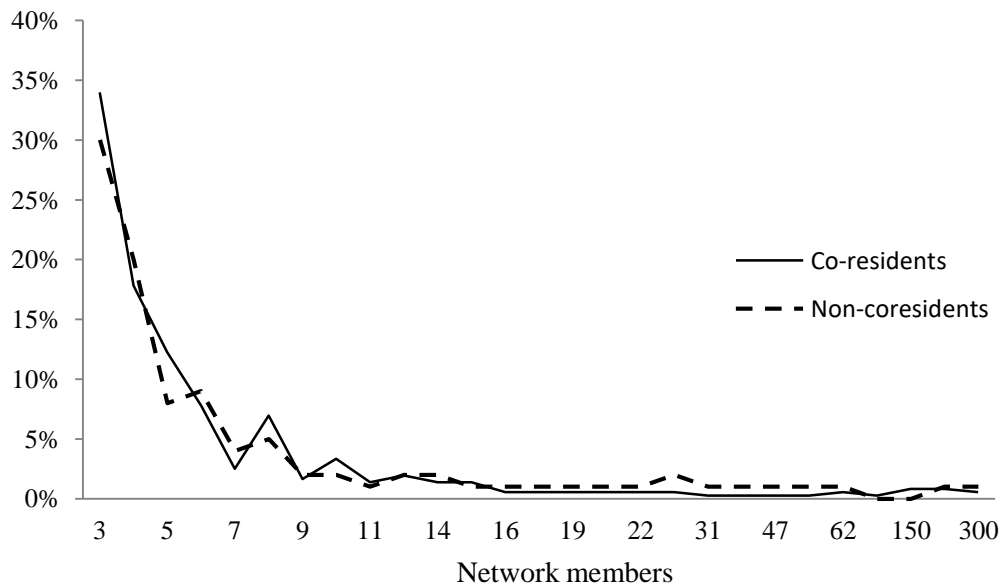
Table 1: Description of data

Variable	Explanation of the variable	Coresidents at Wave III n. obs.1,687		Non-Coresidents at Wave III n.obs 2,221	
		Mean	St.dev.	Mean	St.dev

Wave I variables

Conventional Individual characteristics

Female*	Dummy variable taking value one if the respondent is female.	0.47	0.50	0.57	0.49
White*	Race dummy. "White" is the reference group.	0.72	0.45	0.81	0.41
School grades*	Grade Point Average (GPA) for mathematics, English, history and science. It ranges 1=D or lower, 2=C, 3=B, 4=A.	2.03	0.75	2.94	1.33
Parent education*	Schooling level of the (biological or non-biological) parent who is living with the child, distinguishing between "never went to school", "not graduate from high school", "high school graduate", "graduated from college or a university", "professional training beyond a four-year college", coded as 1 to 5. We considering only the education of the father if both parents are in the household.	3.05	2.23	4.03	1.54
Family income*	Total family income in thousands of dollars, before taxes. It includes income of everybody in the household, and income from welfare benefits, dividends, and all other sources	48.40	52.77	55.68	46.45
Residential area urban	Interviewer's description of the immediate area or street (one block, both sides) where the respondent lives, coded as a dummy taking value 1 if the area is urban-residential only and 0 otherwise (i.e. if the area is rural, suburban, mostly retail, mostly industrial or other type)	0.57	0.49	0.54	0.49

Indicators of social structure of families and parenting

Household size*	Number of people living in the household.	3.55	1.91	2.61	0.99
Dinner spent with parents*	Answer to the question: On how many of the past 7 days was at least one of your parents in the room with you while you ate your evening meal? The answers range between 0 and 7.	5.05	4.23	5.91	3.25
Parental care*	Answer to the question: How much do you think your mother (father) cares about you? , with answers 1=not at all, 2= very little, 3=somewhat, 4=quite a bit, 5= very much. They are averaged between parents.	3.86	2.13	4.66	3.00

Two-parent family*	Dummy taking value one if the respondent lives in a household with two parents (both biological and non biological).	0.67	0.47	0.75	0.43
Married, two-parent family*	Dummy taking value one if the respondent lives in a household with two parents (both biological and non biological) who are married	0.49	0.44	0.63	0.48
Wave III variables					
Age	Grade of student in the current year.	20.98	3.52	21.24	4.43
Married*	Variable taking value one if the respondent is married	0.16	0.37	0.22	0.41
Employed*	Variable taking value one if the respondent is employed	0.70	0.46	0.75	0.43
Network characteristics					
Network size	Number of network members	28.89	25.88	29.02	22.76
Number of nominated friends	Number of friends	2.26	1.88	2.31	1.10

Notes: T-tests for differences in means across groups are performed. Variables marked with * show differences statistically significant at least at the 10% level.

Table 2. Estimation results

	Dep. Var. probability of leaving parental home					
	(1)	(2)	(3)	(4)	(5)	(6)
Peer effects ($\hat{\phi}$)	0.4421**	0.3890**	0.2651**	0.1785***	0.1088**	0.0614
	(0.1998)	(0.1717)	(0.1203)	(0.0654)	(0.0455)	(0.0501)
Standard individual characteristics	no	yes	yes	yes	yes	yes
Additional individual characteristics	no	no	yes	yes	yes	yes
Peers' characteristics	no	no	no	yes	yes	yes
Network fixed effects	no	no	no	no	yes	yes
Individual fixed effects	no	no	no	no	no	yes
N. obs.	3,908	3,908	3,908	3,908	3,908	3,908
N. networks	359	359	359	359	359	359
% of variance explained by peer effects	**35.8%**	**26.5%**	**17.4%**	**9.3%**	**3.6%**	**0%**

Notes. Columns (1) – (5): Maximum likelihood estimation results. Column (6): Bayesian estimation results. Control variables are those listed in Table 1. The additional individual characteristics are "household size", "two parent family", "two married parent family", "dinner with parents" and "parental care". *, **, *** indicate statistical significance at the 10, 5, and 1 percent levels.

Table 3. Bayesian estimation results: Outcome equation and link formation

	Dep. Var.: Probability of leaving parental home		
	(1)	(2)	(3)
	Outcome equation without link formation	Link formation	Outcome equation with link formation
Peer effects (ϕ)	0.1088**		0.0614
	(0.0455)		(0.0501)
Female	0.1574 ***	-0.1654***	0.1672***
	(0.0320)	(0.0402)	(0.0555)
White	0.1405***	-0.5897***	0.1403**
	(0.0402)	(0.059)	(0.0705)
School grade	0.0501***	-0.0689	0.0479*
	(0.0139)	(0.0530)	(0.0267)
Age	0.0235	-0.9347***	0.0245
	(0.0231)	(0.0661)	(0.063)
Employed	0.1511**	-0.1631***	0.1980**
	(0.0721)	(0.0425)	(0.0856)
Married	0.1431 ***	-0.3439 ***	0.1234**
	(0.0441)	(0.0671)	(0.0606)
Residential area urban	0.0985***	-0.0787**	-0.0888 **
	(0.0315)	(0.0369)	(0.0407)
Parent education	0.0798***	-0.0215	0.0811 ***
	(0.0215)	(0.0355)	(0.0433)
Family income (*1000)	0.0998***	-0.0658 *	0.0865**
	(0.0217)	(0.0370)	(0.0472)
Household size	-0.0224*	-0.0574**	-0.0197
	(0.0147)	(0.0273)	(0.0399)
Two-parent family	0.2347***	0.0035	0.2346 ***
	(0.0155)	(0.0202)	(0.0422)
Married, two- parent family	0.0625 **	-0.0495	0.0724***
	(0.0210)	(0.0465)	(0.0274)
Dinner spent with parents	0.1769***	-0.1021	0.1810 ***
	(0.0515)	(0.1050)	(0.0603)
Parental care	0.0226 **	-0.0249*	0.0229**
	(0.0100)	(0.0135)	(0.0124)
Constant		-1.323***	
		(0.0909)	
Unobservables ($\sigma_{s\xi}$)		*0.5896***	*0.0105**
		(0.0201)	*(0.0051)*
Peers'characteristics	yes	-	yes
Network fixed effects	yes	-	yes
Obs	3,908	7,634,287	3,908

Notes. Columns (2) and (3) report the means and the standard deviations of the posterior distributions of the parameters. We draw random samples from each parameter's marginal conditional distribution using Markov Chain Monte Carlo (MCMC) techniques. We let our chain run for 70,000 iterations, discarding the first 10,000 iterations. Ergodicity of the Markov Chain is achieved quite quickly. Control variables are described in Table 1.In columns (2) the regressors are differences in terms of the listed variables between friends. Column (2) reports results on the dyadic model (3), the covariates are differences in terms of the listed characteristics. Columns (1) and (3) show results on model (1)-(2). *, **, *** indicate that zero is not contained in a 90, 95, and 99 percent confidence interval

Table 4. Boomerang kids
-Significant differences-

	Kids non-coresident at wave III and at wave IV		Kids non-coresident at wave III and coresident at wave IV		
	Mean	St. dev	Mean	St. dev	t-test
Female	0.72	0.45	0.43	0.49	0.0000
White	0.79	0.40	0.81	0.41	0.2578
School grades	2.97	1.43	3.01	1.55	0.1102
Age	21.24	4.43	20.61	3.54	0.3279
Employed	0.75	0.43	0.76	0.43	0.4006
Married	0.22	0.41	0.21	0.41	0.5403
Residential area urban	0.43	0.49	0.45	0.67	0.1255
Parent education	3.95	1.65	4.05	1.15	0.2800
Family income	52.66	48.99	60.80	20.40	0.0002
Household size	2.59	0.90	2.50	0.46	0.1706
Two-parent family	0.74	0.44	0.71	0.31	0.1673
Married, two- parent family	0.49	0.44	0.48	0.50	0.2929
Dinner spent with parents	5.50	4.34	5.37	4.38	0.4704
Parental care	4.50	3.50	4.70	3.20	0.4550
Network size	26.01	30.29	25.38	27.34	0.2276
Number of nominated friends	1.98	3.04	1.95	3.05	0.4110
N.obs.	1,629		592		

Notes: T-test for differences in means with unequal variances had been performed. p-values are reported

Figure 3. Distribution of individuals who are not coresidents at Wave III by number of social contacts and residential choices at Wave IV
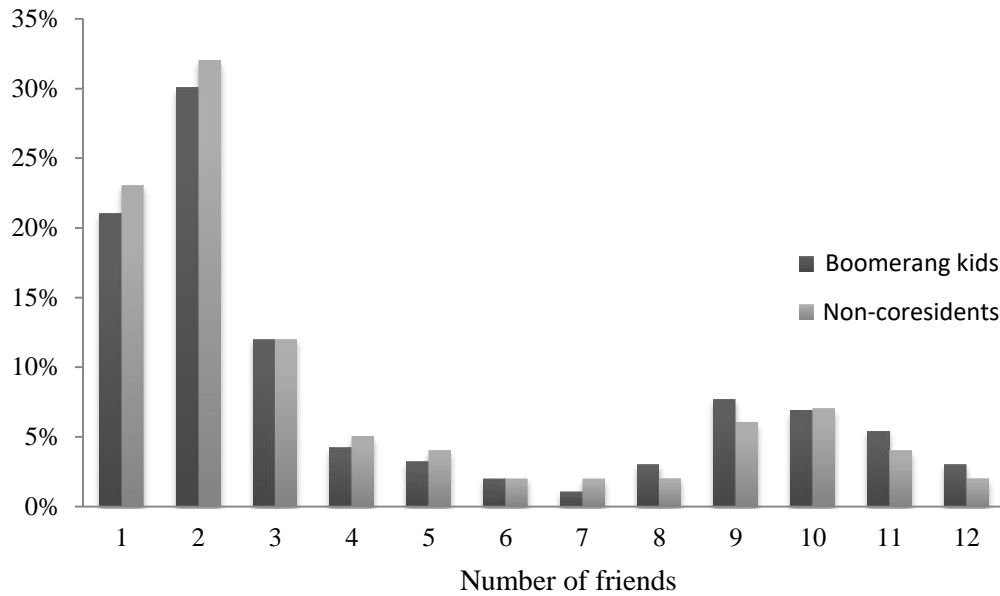
Figure 4. Distribution of networks by network size and residential choices -Sub-sample: individuals who are not coresidents at Wave III-
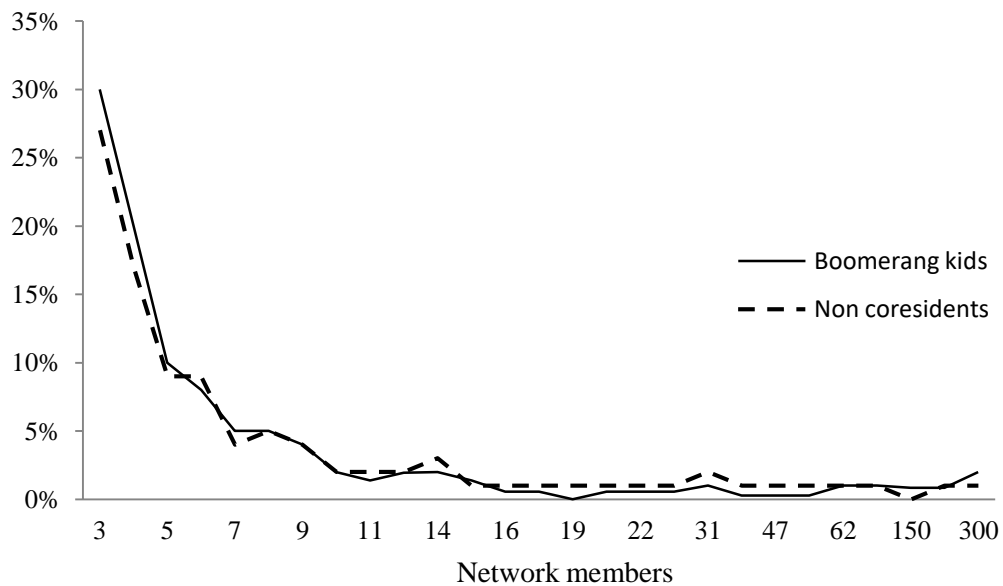
Table 5. Bayesian estimation results: outcome equation and link formation
- Boomerang kids -

	Dep. Var.: Probability of leaving parental home		
	(1)	(2)	(3)
	Outcome equation without link formation	Link formation	Outcome equation with link formation
Peer effects (φ)	0.1569**		0.1644 **
	(0.0501)		(0.0750)
Female	0.1667 ***	-0.1599***	0.1721***
	(0.0320)	(0.0439)	(0.0565)
White	-0.1540***	0.5555***	0.1495**
	(0.0442)	(0.0777)	(0.0723)
School grade	0.0650***	0.0568	0.0697**
	(0.0213)	(0.0630)	(0.0303)
Age	0.0263	-0.7347***	0.0274
	(0.0323)	(0.0766)	(0.0422)
Employed	0.1211**	-0.1863***	0.1198**
	(0.0629)	(0.0545)	(0.0600)
Married	0.1054**	-0.3041 ***	0.1153**
	(0.0504)	(0.0699)	(0.0560)
Residential area urban	0.1509**	-0.0978**	0.1408 *
	(0.0733)	(0.0493)	(0.0789)
Parent education	0.0781***	-0.0415	0.0815 ***
	(0.0195)	(0.0432)	(0.0241)
Family income (*1000)	0.0870**	-0.0657 **	0.0904 **
	(0.0417)	(0.0322)	(0.0453)
Household size	-0.0442**	-0.0474	-0.0505**
	(0.0198)	(0.0403)	(0.0253)
Two-parent family	0.2541***	0.0053	0.2613 ***
	(0.0315)	(0.0298)	(0.0442)
Married, two-parent family	0.1062 **	-0.0345	0.1107**
	(0.0521)	(0.0405)	(0.0574)
Dinner spent with parents	0.1769***	-0.0925	0.1810 ***
	(0.0515)	(0.1205)	(0.0603)
Parental care	0.0450**	-0.0255	0.0492**
	(0.0230)	(0.0212)	(0.0254)
Constant		-3.012***	
		(0.5066)	
Unobservables ($\sigma_{s\xi}$)		0.6968***	0.01206**
		(0.0505)	(0.0610)
Peers'characteristics	Yes	-	Yes
Network fixed effects	Yes	-	Yes
Obs	3,908	7,634,287	3,908

Notes.Columns (2) and (3) report the means and the standard deviations of the posterior distributions of the parameters. We draw random samples from each parameter's marginal conditional distribution using Markov Chain Monte Carlo (MCMC) techniques. We let our chain run for 70,000 iterations, discarding the first 10,000 iterations. Ergodicity of the Markov Chain is achieved quite fast. Control variables are described in Table 1.In columns (2) the regressors are differences in terms of the listed variables between friends. Column (2) reports results on the dyadic model (3), the covariates are differences in terms of the listed characteristics. Columns (1) and (3) show results on model (1)-(2). *, **, *** indicate that zero is not contained in a 90, 95, and 99 percent confidence interval

Table 6. Additional evidence: Boomerang kids and self-control

	(1)	(2)	(3)	(4)	(5)
	Dep. Var.: Self-control				
Boomerang kid	-0.0325**	-0.0810**	-0.0021*	-0.0091*	-0.0234**
	(0.0159)	(0.0380)	(0.0011)	(0.0052)	(0.0114)
Individual characteristics	yes	yes	Yes	yes	yes
Peers' characteristics	yes	yes	Yes	yes	yes
Network fixed effects	yes	yes	Yes	yes	yes
N. obs.	3,908	3,908	3,908	3,908	3,908
N. networks	359	359	359	359	359

Notes. OLS estimation results. Self-control is measured in columns (1) – (5) using the following question: (1) "When making decisions, you usually go with your "gut feeling" without thinking too much about the consequences of each alternative?", coded 1= strongly disagree to 5= strongly agree; (2) If you wanted to use birth control, how sure are you that you could stop yourself and use birth control once you were highly aroused or turned on?, coded 1=very unsure to 5=very sure; (3) When you have a problem to solve, one of the first things you do is get as many facts about the problem as possible, coded 1= strongly disagree to 5= strongly agree; (4) When making decisions, you generally use a systematic method for judging and comparing alternatives, coded 1= strongly disagree to 5= strongly agree; (5) How often was the following true during the past week? You had trouble keeping your mind on what you were doing, coded 1= most or all of the time to 4= never or rarely. Control variables are those listed in Table 1 (and used in Tables 3 and 5). *, **, *** indicate statistical significance at the 10, 5, and 1 percent levels.

Table 7. Robustness checks: alternative definition of network links

Dep. Var. Probability of leaving parental home

	(1)	(2)
	All sample	Boomerang kids
Panel (a)		
Directed Networks (outdegree)		
Peer effects ($\hat{\phi}$)	0.0659	0.1605 **
	(0.0610)	(0.0801)
Directed Networks (indegree)		
Peer effects ($\hat{\phi}$)	0.0715	0.1765**
	(0.0599)	(0.0755)
Panel (b)		
Undirected Networks		
Strong ties	0.0777	0.1917**
	(0.0555)	(0.0935)
Weak ties	0.0314	0.0875*
	(0.0266)	(0.0472)
Individual characteristics	yes	Yes
Peers' characteristics	yes	Yes
Network fixed effects	yes	Yes
N. obs.	3,908	3,908
N. networks	359	359

Notes: Bayesian estimation results. Control variables are those listed in Table 1. *, **, *** indicate statistical significance at the 10, 5, and 1 percent levels.