Social Networks in Higher Education: Estimating Spillovers With Experimental Variation

Kevin M. Williams
University of California, Davis
kmwilliams@ucdavis.edu

December 28, 2015

Abstract

Everyone remembers their best friend from college, but few can recall who sat to their left in English 101, hence, not all peers are created equal. While traditional studies of peer spillovers assume students interact in groups, I model peer spillovers using a social network approach where students are impacted through their closest social connections. This approach implies the effects of spillovers will be concentrated among small social cliques, rather than distributed evenly throughout peer groups. Data come from two cohorts of students at The United States Air Force Academy, where squadrons create distinct social groups and the grading structure allows for uniform measurement of achievement. To account for selection into friendship, I use a novel instrument constructed from students being experimentally assigned to one social network in their freshman year and then randomly re-assigned sophomore year. Data on social ties come from a survey where students identify their friends and study partners. A novel imputation technique is used to account for missing survey data. Results show that social connections, especially study partners, create positive achievement spillovers. My ability to isolate the role of study partners is unique to the literature and offers a concrete mechanism for how peer effects are disseminated among students.

I thank Scott Carrell, Hilary Hoynes, Colin Cameron and Marianne Bitler for excellent advice and guidance. Further thanks to Marianne Page, Kevin Gee and seminar participants at UC Davis, University of Connecticut and the 2015 All-California Labor Conference. The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the USAF, DoD, or the U.S. Government. This article was completed with data that is managed in collaboration with Lt. Col. Scott Carrell and Jim West.

1 Introduction

High-quality peers are seen as a driver of academic success. Spillovers from social interactions create welfare-increasing externalities for students and schools alike (Hoxby 2000). This is one reason why every year students spend countless hours in an effort to attend the best possible college. However, the nature of these effects is far from adequately understood. Existing ways of studying spillovers are limited for a variety of reasons. In this paper, I study spillovers on achievement through social networks, leveraging data where students are randomly assigned to peer groups across multiple periods, to understand more precisely what happens between students inside of lecture halls, libraries and dorms to produce these social spillovers.

The causes of spillovers have implications for administrators and instructors. Consider an administrator deciding whether or not to spend the money to reduce a lecture from a single 150-student class to three 50-student ones. If spillovers are created by "role-model" students who rub-off on their peers and raise their achievement, three separate classes may limit access to the best role-models and actually decrease spillovers. On the other hand, if spillovers occur when students interact in one-on-one settings, smaller classes may facilitate collaboration and friendships that would not otherwise exist. To arrive at the best decision, an administrator needs to know not only that spillovers exist, but also understand their mechanisms.

The standard approach to studying peer effects has been to use the linear-in-means model. Hoxby (2000), Sacerdote (2001) and Carrell et al. (2009) have used this approach and their findings suggest that positive spillovers among peers do exist-an individual's expected performance will improve if placed in a higher-caliber peer group. However, this model has limitations. While it can diagnose the presence of peer effects, it cannot isolate the mechanisms creating them.

There are two drawbacks to the linear-in-means model. First is the assumption that individuals interact equally in groups (e.g. a lecture, a cohort or a dorm). In the model, an individual's outcome is affected by the mean aptitude of his peers. All peers are weighted equally, meaning an individual is assumed to be equally influenced by all other members of his group. While the assumption is generally necessitated by data constraints, all peers are not created equal. The actions of an individual's close social ties will have a larger impact on him relative to other peers.[1]

Second, the linear-in-means model cannot distinguish between two broad categories of peer effects: contemporaneous effects and contextual effects. Contemporaneous effects are particularly important because they directly create a social multiplier effect, whereas contextual effects do so only indirectly.[2] Further, an increase in contemporaneous effects will increase overall social welfare, while changes due to contextual effects may be zero-sum.[3] The inability of the linear-in-means model to separate these two channels has forced researchers to focus on the existence of peer effects rather than on their mechanisms.

In this paper, I use the social connections that form within groups to overcome both shortcomings of the linear-in-means model. I observe students who were randomly assigned to peer groups in two separate time periods as well as information on student's closest social connections (obtained via survey). The randomization of students controls for selection into groups. This approach allows me isolate two specific mechanisms that potentially create spillovers, friends and study partners. Instead of needing to assume that all members of a peer group exert equal influence, I make the more realistic assumption that these within-group study partners or friends drive peer effects. An individual is influenced by the mean achievement of his connections. The combination of random assignment of students to peer groups over multiple periods along with knowledge of their within-group social networks allows me to separate and causally identify contemporaneous

[1]Haynie (2001) and Duncan et al. (2001) discuss how an individual's closest connections disproportionately shape their educational experience relative to other individuals who happen to be in the same peer group.

[2]Note, these are also called endogenous and exogenous effects, respectively. As defined by Manski (1993), contemporaneous effects represent an individual's performance changing due to the efforts of their peers and contextual effects represent how an individual's performance is impacted by composition of their peer group. Social multipliers are discussed in Section 2. In short, the social multiplier describes the phenomena where one individual increasing own effort inspires some peers to increase their effort, which in turn may inspire more peers, etc. Students responding the effort of their peers is a contemporaneous effect, thus contemporaneous effects directly imply a social multiplier. A contextual effect occurs when an individual responds to his environment with an increase or decrease in effort. The resulting change in effort will have a multiplicative effect, but through contemporaneous means. The initial change in effort in response to context is a one-time change, thus contextual effects only indirectly cause social multipliers.

[3]If contextual effects are linear, then they are zero-sum. The benefit of moving an individual to a new group will be offset by the cost of him leaving his old group. If they are non-linear, then there is potential for overall welfare increases, see Carrell et al. (2013).

and contextual peer effects.

Although peers can be randomly assigned, the exact nature of a social connection between two students always reflects their own choice. This makes endogenous selection of connections a first-order concern when using social network data. Bramoulle, Djebbari & Fortin (2009), Lin (2010) and De Giorgi et al. (2010) has shown how to leverage knowledge of within-group social networks and account for the endogeneity of students choosing their connections using an instrumental variables approach.[4] With an appropriate instrument, the variation in connections among students can be used to separate the contemporaneous and contextual channels of peer effects. My setting differs from previous studies because I observe individuals who were randomly assigned to peer groups over two periods. This double randomization offers a more credibly exogenous instrument than prior attempts, which needed to rely on strong spatial assumptions.

Data come from two cohorts of students at the United States Air Force Academy (USAFA). I measure spillovers in achievement using GPA as the outcome. At USAFA, students in their first two years take a common set of core courses, each of which is taught and graded in a standardized fashion across all sections of the course. This makes GPA a reliable measure of achievement. USAFA students are organized into forty distinct squadrons that make up their core social unit. Students board, dine and study with other members of their squadron. These squadrons make up my distinct peer groups. Social network data of the connections that exist among USAFA students was obtained via survey. On the survey, students listed their closest friends and study partners from their sophomore year. This is the first time, to my knowledge, that information on student's study partners, separate from friendships, has been used in estimating spillovers.

Surveys were completed by a subset of students and to account for unobserved data, I modify the Chandrasekhar & Jackson (2012) model of network imputation to simulate fully observed social networks. My modification adjusts for selection into survey response and for dimensionality. Full details of the imputation, including tests of validity are discussed in the appendices.

[4] Bramoulle, Djebbari & Fortin (2009) formally refers to their approach as instrumental variables while the others use spatial autoregressive models that, in principle, apply a very similar identification strategy.

The observed cohorts of students were subjected to experimental "social-engineering" during their freshman years, described in Carrell et al. (2013). The authors attempted to design squadrons to optimize peer spillovers among students. The experiment did not go according to plan, actually harming some of the students it was designed to help, but it did create exogenous shocks to student achievement. Before their sophomore year, students were randomly *re-assigned* to new squadrons and forced to make new social connections. This feature of USAFA is key to identification. Typically when a social network is observed over multiple periods, the later network is heavily dependent on the prior, but not in this case. The experiment along with the re-assignment of students allows me to use the freshman-year experimental statuses of a student's sophomore friends as an instrument for the friends' sophomore year performance.

Results show that the contemporaneous channel of peer effects dominates the contextual one. This concurs with findings by Bramoulle, Djebbari & Fortin (2009) and Lin (2010). However, my point estimates of the contemporaneous effects, 0.11 are $50\% - 75\%$ smaller than theirs (0.40 and .27, respectively). I show that the discrepancy in estimates is not simply due to USAFA students being different from previously studied populations. When I estimate a comparable model (i.e. one which does not utilize experimental variation as an instrument), I find higher point estimates for contemporaneous effects. This suggests that previous attempts may not have been able to fully account for the endogeneity of students selecting into friendships.

I show that contemporaneous effects are driven by study partners, rather than friends. This is an important finding, since it provides a single, isolated channel that creates positive peer effects. Positive spillovers are created outside of the classroom when students collaborate together on their work. We now have a better idea of how peer effects are created within a group. There does not appear to be a role-model effect, where the mere presence of high-ability peers leads to their talent rubbing off on others. Instead, it is the one-on-one connections students make that are academic in nature that create these beneficial spillovers.

When students actively study together, they all seem to benefit. While intuitive, this phenomenon is challenging to causally identify and has interesting ramifications for how we think about peer effects. The beneficiaries of spillovers are likely to be concentrated among small social cliques of students within the larger peer group. In this case, a rising tide will not lift all boats. Simply placing a student into a group with higher-ability peers will not improve his performance. Social connections with other students are needed.

This means that the *opportunity* to make connections is important when trying to maximize contemporaneous effects. Two individuals can never be assigned to be friends, but there are certain patterns that govern links among students. I present suggestive evidence that students at USAFA who were randomly assigned to peer groups with a higher number of expected connections experienced stronger contemporaneous effects.[5]

Section 2 provides an overview of social networks and shows how the interpretation of contemporaneous effects differs between linear-in-means and social network models. Section 3 reviews the traditional linear-in-means peer effects model, terminology, and channels of peer influence before detailing the social network model used in this paper. Section 4 describes the data and the experimental sorting of students used as an instrument. Section 5 presents results. Section 6 concludes with policy discussion and thoughts on future research.

2 Social Networks in Education

Social networks are a tool for measuring the complex web of friendships, relationships and other connections that form among groups of individuals. They are becoming an increasingly common empirical tool, being applied everywhere from labor to crime to development economics.[6]

[5]This relates to Fletcher et al. (2013) and Bramoulle, Rogers, Currarini, Jackson & Pin (2009) discussions on homophily and friendship supply. The probability of student connections can be estimated using observables such as gender, race and athletics. To calculate an indivdual's expected connections, I sum the probabilities that he is friends with each other member of his squadron. As an example, gender is the strongest predictor of social ties. A femal who is in a squadron with 8 other females will likely have a higher number of expected friends than one who is a squadron with only 4 other females. It should be noted that most of the work on patterns of connection-formation has focused on friendship formation, rather than study partner formation.

[6]See Zax & Reese (2002), Liu et al. (2012) and Chandrasekhar & Lewis (2011), respectively

Research from sociology supports the idea that individuals are dependent on their closest social ties for receiving and transmitting new information (Haynie 2001). It is natural to assume that, among students, social networks play an important role for how peer effects manifest within a school. This section will discuss how social networks are defined and interpreted. Identification of models using social networks will be discussed in Section 3.

A few terms that have broad meaning are used more narrowly in this paper. A *peer* is any other member of their observed educational group (e.g. a lecture, cohort or dorm). In contrast, a *connection* is a peer with whom the student has an explicit relationship such as a friend or study partner. By this measure, a student's group of connections will be a subset of their peers. The *social network* of a peer group is a representation of all the connections within that group. When using social network analysis, individuals in the same peer group are assumed to have distinct, but overlapping reference groups, based on their connections.

In attempts to model a social network, previous work has used multiple different definitions of what constitutes a peer group and a connection among students. Peer groups have been assigned as roommates, dorms, squadrons or cohorts, to name a few.[7] Estimates of peer effects among students have been shown to be sensitive to choice of peer group, see Carrell et al. (2009). The sensitivity of results to choice of peer group supports the idea that peer effects are not in fact created by role-model students, who peers observe from afar and gain inspiration. Rather, peer effects occur among students with explicit social ties and measures of peer groups are noisy proxies for a student's social circle.

A connection between two students has also been defined multiple ways. De Giorgi et al. (2010) considers two students connected if they have sufficient overlap in their randomly-assigned lectures. While this measure does not contain any knowledge of student's actual relationships, it has the advantage of creating a social network where students are randomly assigned their friends. Marmaros & Sacerdote (2006) looks at the volume of e-mail among students to proxy for friend-

[7] See Sacerdote (2001), Foster (2006), Carrell et al. (2009), Bifulco et al. (2011) respectively.

ships. The most popular way to assign a connection among students is through use of survey data. Bramoulle, Djebbari & Fortin (2009), Calvó-Armengol et al. (2009) and Lin (2010) are just a few of the papers who rely on student's survey responses to model their social network. This will be the definition used in this paper as well. Survey data is precise in its interpretation, which makes it appealing for understanding the underlying mechanisms of peer effects. If a student lists another student as a common study partner, say, we have a much better signal of what their connection entails compared to two students with lots of e-mail or course overlap. This means that when we use survey information on connections to estimate peer effects, results will isolate a narrower set of possibilities for what is occurring among students to drive these effects.

Social networks are formally represented by a social graph (adjacency) matrix. For a group of students size n, the social graph g is $n \times n$. $g_{ij} = 0$ if students i and j are not connected. $g_{ij} = \frac{1}{n_i}$ if they are connected, with n_i representing i's total number of connections.[8] Multiple groups of students can be arranged in a block-diagonal matrix, G. G represents the distinct social networks across for multiple groups of students and will be used going forward.

Connections are considered *undirected* if $G_{ij} \neq 0$ implies $G_{ji} \neq 0$. Undirected connections are assumed to be reciprocal. *Directed* friendships do not imply reciprocation. Which definition to use depends on the networks, in the case of students I feel it is best to assume undirected friendships. A student listing another as a friend should imply that those two students spend some amount of time together. The listed student may not consider the listing student a close friend, but still spends time with them and potentially could obtain the sort of spillovers I am interested in estimating. Robustness checks will consider directed networks.

A key difference between social network analysis and traditional linear-in-means models is how peer effects are interpreted. Figure 1 illustrates these differences. Both panels present a fictional six-student class. Students are identified by number in the top left figure. The top row represents the linear-in-means model, whereby all students exert equal influence on each other, represented

[8] Some definitions of a social graph will set $g_{ij} = 1$ for all connections. For my empirical purposes, this definition would give more weight to students who have more friends

by all students being connected. Under the linear-in-means model, own performance is affected by *peer* means. The bottom row gives the social network of the class, with lines indicating connections between two students. In a social network model, own performance is affected by the *friend* mean of a student.

Assume that there is a contemporaneous effect of 0.25. This implies that if the relevant mean performance rises by one unit, there will be a spillover onto own performance (initially) of 0.25. The figures in the right column show the predicted change in performance due to a one-unit increase in student-1's performance (e.g. outside tutoring). In the top row, the linear-in-means approach, all students have five connections so this shock raises *peer* mean performance for students 2-6 by 0.2. The result is an $0.2 * 0.25 = 0.05$ initial spillover effect for each of these students. The 0.05 rise for students 2-6 causes a secondary, indirect, spillover for all students, since they have once again seen peer mean performance rise. The equilibrium values are shown. Students 2-6 all receive an equal expected spillover of 0.067 from the shock to student 1's performance.

In the bottom row, the social network model, spillovers act in a similar fashion, but only operate through connected students. Student-2, for example, has only one connection, student-1, and so the shock to raises student-2's *friend* mean performance by one-unit. Thus student two receives an initial spillover of 0.25. Student-3, who has two connections, one of which is student-1, would receive and initial spillover of 0.125. Secondary and tertiary indirect spillovers exist as before. Using the social network approach, the precise equilibrium magnitude of the spillover depends on a student's number of connections and their proximity to the initial shock. Student 2, whose only connection is with the student receiving the shock, gains the largest benefit. Student-3 and student-4, also connections of 1 both receive substantial predicted spillovers themselves, but they are lessened due to the fact that they have other connections. Student-5 and student-6, who are two and three degrees removed from student-1, respectively, receives severely decayed, almost negligible, spillovers.

This exemplifies how social network analysis predicts the beneficial spillovers created among peers to be concentrated among small cliques of students. Student 1's three closest friends receive an average benefit of 0.137, twice as high as in the linear-in-means model. The students who are not friends with student 1 receive a greatly decayed benefit - less than a fifth of the size of the linear-in-means. Despite the original shock being a one-unit increase, student 1's predicted change is greater due to the spillovers he receives from his friends and classmates improving their performance.[9]

3 Model

Peer effects models seek to estimate the role that an individual's peers play in determining outcomes. Section 3.1 starts by walking through the Manski (1993) linear-in-means model of peer effects. I define contemporaneous effects, contextual effects and correlated effects. I discuss how the assumptions of the model and identification challenges make pinning down the mechanisms that drive peer effects difficult. Then, in Section 3.2, I show how incorporating social-network information builds on the linear-in-means model and is able to separately identify the various channels of peer effects. This creates the ability to isolate specific mechanisms. Lastly, Section 3.3 discusses the adjustments that are necessary to account for the data imputation needed in my setting.

3.1 Linear In Means Model

We want to understand how a student's performance is impacted by those around them. There are three channels of interest - contemporaneous effects, contextual effects and correlated effects, represented in the following model:

$$Y_{ic} = \overline{Y}_{c,-i}\alpha_1 + \overline{X}_{c,-i}\alpha_2 + X_{ic}\alpha_3 + \gamma_c + \epsilon_{ic} \tag{1}$$

[9] The greater total magnitude of the social network case (combined 1.56 increase) relative to linear-in-means (combined 1.352) is not important. As group size grows, the magnitude of both total improvements will quickly converge to 1.33, the value of the social multiplier. For a contemporaneous effect size of β the social multiplier is given by $\frac{1}{1-\beta}$. See Carrell et al. (2013).

Y_{ic} is the performance of observed student i in classroom c[10]. X_{ic} represents predetermined characteristics and measures of ability (e.g. gender, race, SAT scores). $\overline{Y}_{c,-i}$ is the mean classroom peer performance of all students except i. $\overline{X}_{c,-i}$ represents mean peer background ability. γ_c are classroom fixed effects.

α_1 represents contemporaneous (a.k.a. endogenous) peer effects, the main parameter of interest. It describes how an agent's actions change according to the actions of their peers. For example, a student who receives an external shock to their performance (e.g. outside tutoring) begins studying hard, raising his achievement. Other students observe this and improve their own achievement. Contemporaneous effects are particularly important because they result in a social multiplier.[11] One student's efforts can compound and benefit all his peers.

There is a broad range of real-world mechanisms that could result in contemporaneous effects. An insightful question during a lecture, a study group among friends, or students conspiring to cheat on a test are a few examples of actions that would be considered contemporaneous effects. This means that identifying α_1 is not the same as isolating actual within-classroom mechanisms that create these effects. How to interpret α_1 depends on the particular setting and variation used.

The vector α_2 measures contextual (a.k.a. exogenous) peer effects. It describes how student performance is affected by the characteristics of their peers, independent of performance. For example, students may change their behavior based on the ratio of boys to girls in a classroom, regardless behavior. Alternatively, a student who wins a school lottery and ends up at a better school may feel more pressure to work hard simply because of being surrounded by higher-caliber peers. When linearity is assumed, contextual effects represent a zero-sum scenario whereby moving a student from one peer group to another will harm one group as much as it helps the other. The zero-sum nature of contextual effects makes them a less appealing target for policy. Within this

[10] Here I'm using the generic term "classroom," which depending on the setting could be any educational unit such as a homeroom, a cohort or, in my case, a squadron.

[11] If contemporaneous effects do not exist ($\alpha_1 = 0$) a student's improvement of δ results in a $\frac{\delta}{N}$ rise is group mean. With positive contemporaneous effects, $\alpha_1 > 0$, the same increase causes an expected rise in group mean performance of $\frac{\delta}{N(1-\alpha_1)}$. The value of the social multiplier is defined as $\frac{1}{1-\alpha_1}$.

model, finding a way to increase α_1, so that students respond more strongly to the achievement of their peers, offers a clearer path towards making a meaningful gain in overall social welfare.

γ_c are correlated effects. They represent fixed differences across classrooms. Depending on the setting of the data, these differences could include selection into classrooms, teacher quality, or shocks to classrooms (e.g. poor test conditions on the day of an exam). Correlated effects are not peer effects, but they represent alternate causes of why different classrooms may diverge in mean achievement. Thus correlated effects are confounders for identifying contemporaneous and contextual peer effects and must be accounted for.

The assumption of Equation 1 is that student attainment responds to peer means. This gives all peers in a classroom equal weight, predicting that a student's attainment will not depend on the relationships built inside the classroom. This assumption is not likely to accurately reflect how social interactions and spillovers occur within a classroom. Obviously, it is an assumption that is more often made by necessity (data on within-group social interactions is relatively rare compared to data on how students are grouped), rather than by choice.

The linear-in-means model, as written in Equation 1 is not identified. First, contemporaneous effects and correlated effects, α_1 and γ_c respectively, are co-mingled. Imagine two classrooms, each containing the same number of students with equal values of $\overline{X}_c$ where one classroom uniformly out performs the other. The observed data will look similar whether the performance difference was due to a high quality teacher or to positive contemporaneous effects caused by an unobserved shock to one student's effort. Second there is an issue of simultaneous causation, the "reflection-problem." Some students may be providers of peer effects and others recipients. A given student's Y_{ic} may be causing $\overline{Y}_{c,-i}$ to rise rather than the other way around. Further, Y_{ic} is a function of X_{ic}, which implies $\overline{Y}_{c,-i}$, the contemporaneous effects channel, will be a function of $\overline{X}_{c,-i}$, contextual effects. Identifying variation in this model comes from comparing attainment across classrooms.

The identification challenges and assumptions of the linear-in-means model make it inadequate for estimating causal mechanisms that drive peer spillovers. Estimating Equation 1 as written produces biased estimates of α_1 and α_2. Hoxby (2000), Sacerdote (2001) and Carrell et al. (2009) have successfully used idiosyncratic or random variation in peer group assignment to estimate a model of the following general form:

$$Y_{ic} = \overline{X}_{c,-i}\beta_1 + X_{ic}\beta_2 + \bar{\epsilon}_{ic} \tag{2}$$

The idiosyncratic or random variation allows for the plausible assumption that there is no selection into peer groups and that students face a common environment so that γ_c is either not present or small. β_1 represents a combination of α_1, α_2 and α_3 from Equation 1. It represents the presence of either contemporaneous or contextual peer effects.[12] To account for simultaneous causality and the interdependence of $\overline{X}_{c,-i}$ and $\overline{Y}_{c,-i}$, estimation is forced to combine the two primary channels of peer effects. Thus, the linear-in-means model, in certain situations, is adequate at identifying the presence of peer effects, but struggles to pin down causal mechanisms driving them.

3.2 Social Network Models

Social network models generalize the linear-in-means framework. Rather than interacting in groups, students are assumed to interact through their connections with other group members. Under this assumption, there is significant within-group variation in mean performance and characteristics of students' immediate connections. However, the researcher also must account for selection into connections on the part of students. This section will discuss how, with a valid instrument, incorporating social networks into the model will allow for causal and separate identification of contemporaneous and contextual effects.

The structure of my social network model of peer effects is:

[12] Carrell et al. (2013) shows how to get this result by solving to remove $\overline{Y}_{c,-i}$ from the right-hand-side and taking the limit as the size of a classroom grows. In terms of the original parameters you have: $\beta_1 = \alpha_3, \beta_2 = \frac{\alpha_2+\alpha_1\alpha_3}{1-\alpha_1}, \bar{\epsilon}_{ic} = \frac{\gamma_c}{1-\alpha_1} + \epsilon_{ic}$

$$Y_c = \alpha_1 GY_c + \alpha_2 GX_c + \alpha_3 X_c + \gamma_c + \epsilon_c \tag{3}$$

This takes the same basic form as Equation 1. α_1, α_2, α_3 and γ_c are analogous to the linear-in-means model. The difference is the inclusion of G, the matrix of the social graph. Defined in Section 2, G represents all of the connections between students. GY_c is a vector of friend-mean achievement and GX_c represents friend-mean background characteristics. If everyone in a classroom were connected, then $GX_{ic} = \overline{X}_{c,-i}$, $GY_{ic} = \overline{Y}_{c,-i}$ and Equation 3 would be identical to Equation 1.

The inclusion of G, the social network of a peer group, changes the assumptions of who students are affecting and being affected by within their peer group. Now, students are interacting with their small group of closest friends, rather than all of their peers. This means that students in the same peer group are exposed to varying, but overlapping set of interactions. The social network creates a web-like structure of interactions. The additional layer of variation provided by G is ultimately what will be used to separately identify contemporaneous and contextual effects.

Incorporating social network information into the model does not immediately remove the endogeneity concerns that plague Equation 1. In fact, it adds in an additional hurdle - endogenous selection of who to associate with. Estimating Equation 3 as an OLS model would provide upward-biased estimates of α_1 and α_2 as the parameters would capture the propensity of students to befriend others of similar ability.

The first hurdle, correlated effects, can be addressed by de-meaning Y_c and GY_c at the classroom level, equivalent to having a classroom fixed effect. The assumption that students are only exposed to their direct connections means there is substantial within-class variation in GY_{ic}.[13] In the linear-in-means model, correlated effects are a concern because the model is identified off of across-classroom variation in performance. By de-meaning all variables at the classroom level, across-classroom variation in attainment is no longer part of the identification, but the result is that

[13] In the linear-in-means model, there is minimal variation in $\overline{Y}_{c,-i}$, which is only created by excluding one's self. The variation moves towards zero in the limit.

correlated effects of classrooms no longer contaminate estimates of contemporaneous effects. This is perhaps the largest difference between the social network and linear-in-means approaches to peer effects. Rather than using variation in mean-departures of entire peer groups, Equation 3 uses within-classroom variation in student performance to identify the effects of social connections.

The issues of selection into friendship and simultaneous causality are addressed by instrumenting for GY_{ic}. Following Bramoulle, Djebbari & Fortin (2009), a valid instrument for GY_{ic} will result in $G\widehat{Y_{ic}}$, a predicted value of mean performance among i's connections, that is independent of any influence i had on their friends.[14]

Previous estimates of social network models have used the background characteristics of those two-degrees removed from a student (i.e. their friends-of-friends), G^2X as instruments for mean friend performance GY. The argument for the validity of friends-of-friends characteristics is akin to a time series model where period t will be instrumented by a period $t-1$ or to a spatial model where an agent's neighbor's performance will be instrumented with their neighbor's neighbors. If student i only interacts with their established connections, then those students who interact with i's connections, but not with i, will only influence i insofar as they impact i's social circle. However, classrooms and social networks do not exist in rigid time or space. Using friends-of-friends as an instrument will violate the exclusion restriction if i directly interacts with students outside his immediate circle.

The assumption that friends dictate influences in classrooms is a better one than that of linear-in-means models, where students face equal influence from all peers. It is not, though, a perfect representation of the social dynamic of a classroom. Thus, to the degree that the core assumption does not hold, the problems facing the exclusion restriction on G^2X as a valid instrument are twofold.

[14] Bramoulle, Djebbari & Fortin (2009) show that the existence of an intransitive triad (a student who is at least two-degrees removed from another student) is required to have sufficient variation in GY_{ic}. While I will not test for this formally, it will always be the case that this is true in groups of more than just a few students. USAFA squadrons are large and friendships are sufficiently sparse that it is never the case that all students in a squadron are friends with each other.

A contribution of my paper is to bring a new, experimentally-driven instrument to the estimation of this model. As described in Section 4, students at USAFA were subjected to an experiment their freshman year that created an exogenous impact on their grades. I estimate a peer effects model in their sophomore year, one year after the experiment and after students were randomly reassigned to new squadrons. If student i is connected to students j and k sophomore year, then j's and k's experimental status freshman year will be used as an instrument for GY_{ic}. This instrument is correlated with j and k's performance (which make up GY_{ic}) and the only impact on student i will be through the sophomore performance of j and k. This instrument will create unbiased estimations of α_1, contemporaneous effects. The first stage equation is:

$$Y_{jc} = \zeta_1 Status_j + \zeta_2 \overline{X}_{c,-j} + \zeta_3 X_{jc} + \upsilon_{jc} \tag{4}$$

$Status_j$ is the set of instruments representing the experimental treatment status of a student, describe fully below. $Status_j$ is distinguished from other characteristics X in that it is experimentally assigned. However, it is similar to the other characteristics in that it is treated as an exogenous, pre-determined characteristic of each student. In this sense, my approach is similar to those of Bramoulle, Djebbari & Fortin (2009) and De Giorgi et al. (2010). In those papers, the authors control for an individual's characteristics directly, then use characteristics of other students (in their cases, friends-of-friends) as instruments. I likewise directly control for a student's own experimental $Status_i$, then use the experimental $Status_j$ of other students as an instrument. In my unique setting, experimental status was assigned during freshman year and is uncorrelated with sophomore year peer group assignment and thus can be thought of as an exogenous background characteristics of each student in their sophomore years.

In the first stage, I control for a student's freshman-year *peer* group, $\overline{X}_{c,-j}$, rather than their group of social ties. A student's peers are randomly assigned and so this avoids including selection into social connections in my first stage.

Because the variable being instrumented is an average, I do not use two-stage least squares.

Instead, in what can be thought of as my first stage, I instrument for Y_c at the individual level, then construct $G\widehat{Y_c}$ for each student. For example, say that student i has two friends, a and b. The ith row of GY_c is equal to the average of a's and b's outcome. A two-stage least squares approach would require instrumenting directly for GY_c, thus instrumenting for i's average friend performance. Instead, I instrument for each student individually. In this example, I obtain a distinct value $\widehat{Y}_{ac}$, student a's predicted performance and $\widehat{Y}_{bc}$, student b's. The ith row of $\widehat{Y_c}$ will be the average of $\widehat{Y}_{ac}$ and $\widehat{Y}_{bc}$.

Incorporating the instrument results in the following, identified and estimable model:

$$Y_c = \alpha_1 G\widehat{Y_c} + \alpha_2 GX_c + \alpha_3 X_c + \epsilon_c \tag{5}$$

3.3 Account for Missing Surveys

Estimation of Equation 5 assumes knowledge of friendships for every observation. As will be discussed in Section 4, friendship nominations data at USAFA were obtained from a subset of students so connections for non-surveyed individuals need to be estimated. To account for partial sampling, I will utilize a model of network reconstruction from Chandrasekhar & Jackson (2012) that allows me to consistently estimate the full social network. The authors outline a subgraph generation model that allows for consistent network reconstruction in empirical settings.[15]. Along with Chandrasekhar & Lewis (2011), the authors show how relatively tight and consistent parameter estimates are possible from partial network data, even with the majority of social ties unreported.

When I account for missing data, the resulting, estimated, model is:

$$Y_{ic} = \alpha_1 G^m \widehat{Y}_{ic} + \alpha_2 G^m X_c + \alpha_3 X_{ic} + \epsilon_{ic} \tag{6}$$

[15]Their model is an evolution of exponential random graph models and strategic network formation models. See Christakis et al. (2010) and Jackson (2010) for a full history of these models

Student characteristics and performance, X and Y respectively are observed for all students and are unchanged from above. The difference is, G^m, which represents the mth reconstructed social network. In total, I impute the network in 250 distinct simulations. Results will be reported as distributions of point-estimates over the independent repetitions.

The network reconstruction is a crucial element to my estimation. Appendix A details the full process of how network data is imputed, including modification made to Chandrasekhar & Jackson (2012) in order to account for non-random selection into survey taking. Appendix B provides tests the validity of the imputation model. Using a separate data set with a fully observed social network, I replicate the model of peer effects from Lin (2010). Then, I simulate my partial survey response rate, using only a fraction of the data set's friendship nominations. With this subset, I impute missing observations and re-estimate the model. I show that the reconstruction performs very well, with all coefficients failing to reject the hypothesis that they are different from their true values.

4 Data

4.1 Institutional Setting

The analysis is performed using data from The United States Air Force Academy (USAFA). USAFA is an undergraduate institution with total enrollment of approximately 4,400 students. The period studied covers students from the graduating classes of 2011 and 2012, during their freshman and sophomore years, 2411 students in total. Students are geographically diverse and the average combined SAT Math and Verbal for the cohorts was 1,300.[16] Data combines administrative data on student's backgrounds and attainment, experimental sorting of students to peer groups and survey data on students social ties.

My outcome variable is student GPA, measured on a 4-point scale. Students at USAFA face a large set of core courses in engineering, humanities, social sciences, basic sciences, military

[16]See Carrell et al. (2009) for further description of courses, admissions and finances.

studies and physical education. Freshman and sophomore years are largely spent taking required core courses.[17] Courses are taught in small sections with around 20 students per section and a required course may have eight or more sections per semester. Common tests at USAFA are given during a common testing period while syllabuses and grades are standardized across all sections of a course. Schedules and professors are assigned without any input from the affected students. These features make GPA a consistent measure of performance.

Data from admissions records provides a rich set of information of student demographics and pre-determined measures of ability. This includes gender, race, whether or not the student is a recruited athlete, SAT Math score, SAT Verbal score, academic entry score, fitness entry score and leadership entry score. The academic entry score is a weighted mixture of a student's high school GPA, their class rank and the quality of their high school. Fitness score is based on fitness tests completed their senior year of high school and leadership score measures the value of their high school and community activities.

Attrition is responsible for an 11% reduction from initial sample of 2411 students to the analyzed sample of 2144 students. Analysis is focused on sophomore year attainment. Students who began at USAFA, but left or dropped out before earning a sophomore year GPA are dropped from the sample. There is no identifier for when students left USAFA. There is some attrition during basic training, a three-week course for freshman that meets before the academic school year begins while others dropped out during freshman or sophomore year (before grades were given out). If a dropped student has an assigned freshman squadron and freshman GPA, but no sophomore GPA, their presence is included when constructing variables based on peer means in freshman year.

Student summary statistics are provided in Column 1 of Table 1. 21% of the sample is female, 5% is black, 8% is Hispanic and 8% is Asian. 23% are recruited athletes. The mean SAT math and SAT verbal scores are, respectively 659 and 633. 17% of students attended a college preparatory school. Entry scores are normalized within each cohort. The average sophomore GPA of these

[17] An exception is foreign languages. Students must take foreign language classes, but can choose their preferred language

students is 2.84. The rows labeled "High Verbal", "Other" and "Predicted Low" relate to the experimental treatment students were subjected to in their freshman year, described below. The make-up of USAFA students is comparable to other selective institutions. The SAT scores of students compare with flagship schools such as UCLA and UNC Chapel Hill, while the heavily-male demographics of USAFA are more similar to other technical institutions such as Georgia Tech and Rensselaer Polytechnic Institute.

The military setting distinguishes USAFA and raises questions of external validity. How comparable are these selective students to other university students in less selective or less distinct settings? Prior research done using USAFA data suggests that the mechanisms driving educational attainment work similarly at USAFA as elsewhere.[18] While not conclusive, there is no clear pattern to suggest that mechanisms impacting academic or social outcomes operate differently within USAFA than elsewhere in higher-education. It seems to be that the features which make USAFA distinct, such as random assignment of course schedules and squadrons as well as a structured school day, do not drastically alter student responses to social and educational stimuli. The structure of USAFA make it an excellent setting for identifying new and important questions within education research.

4.2 Experimental Sorting

The primary social units at USAFA are student squadrons. Upon entering USAFA, students are grouped into one of 40 squadrons. Each group is comprised of approximately 110 students (freshman through seniors). Students of a squadron live in adjacent dorm rooms, dine together, study together, compete in intramural sports together and perform military training together. Squadron assignments are random, conditional on balancing certain demographics. For their first seven months

[18] A few examples. Carrell et al. (2010) use data from USAFA and find that a professor's gender influences student's choices of major. Feld & Zoelitz (2015) studies students at a non-selective university and finds similar impacts. Carrell et al. (2013) experimentally construct social units at USAFA (used as the first stage in this paper) and find that social groups with low-variance among student aptitude perform best. Booij et al. (2015) run a similar experiment, again at a less selective, non-military institution and find results of similar size and strength.

in the academy, freshman students are not allowed to enter the premises of another squadron.

Critical for this study, students are randomly assigned to squadrons twice while at USAFA. They are assigned to one squadron as a freshman and then randomly re-assigned to a new squadron at the start of their sophomore year. The sophomore squadron will present them with a new set of peers and will force the creation of new connections and a new social network.[19] Students remain in their sophomore squadron for the next three years.

The two observed cohorts were subject to experimental treatment in the design of their freshman-year squadrons. Full details of the experimental process can be found in Carrell et al. (2013). The experiment was motivated by a desire to optimally assign peers in order to help bottom tercile students raise academic achievement. Prior cohorts at USAFA had shown that bottom tercile students benefited from having peers with high verbal ability in their squadron. The experiment intervened in USAFA's typical random assignment of students to their freshman squadrons and instead designed the squadrons in such a way as to maximize overlap of these predicted low achievers with peers of high verbal ability.

For the experiment, half the students were randomly assigned to be controls. Control students were placed into squadrons created using USAFA's normal stratified randomization. The other half were subject to treatment. A sorting algorithm was used to place students in a way intended to maximize gains in achievement for bottom tercile students. This resulted in two types of treatment squadrons: bifurcated and homogeneous. Bifurcated squadrons contained mostly predicted bottom terciles students and students with high verbal ability. These squadrons were predicted to have a large, positive impact on the bottom tercile students. Homogeneous squadrons were mostly made up of students who had been assigned to the treatment group, but were neither predicted to be in the bottom tercile nor had high verbal ability. They contained mainly middle-tercile students and had significantly lower variation in pre-determined measures of aptitude.

The experiment produced unexpected results. Conditional on incoming measures of ability,

[19]On average, students have .97 members from their freshman squadron present in their sophomore squadron.

students in the homogeneous treatment squadrons performed the best. In the bifurcated squadrons, top tercile students, who had a higher percentage of top tercile peers than the average control group, performed marginally better. Bottom tercile students in the bifurcated squadrons, who the experiment was designed to help, fared the worst. The experiment caused an exogenous shift in grades for those in both bifurcated and homogeneous treatment squadrons.

A student's experimental status is the instrumental variable used when estimating Equation 5. A student's experimental status in their freshman year is used to instrument for their attainment sophomore year. The experiment had non-monotonic effects on treatment students so a set of indicator variables are used, rather than a single indicator for treatment vs control. The four indicators are *Treatment* $\times$ *Bottom Tercile*, *Treatment* $\times$ *Middle Tercile*, *Treatment* $\times$ *Top Tercile* and *Control* (omitted category). These indicators capture how the experiment differentially impacted students of different predicted ability levels. Since students had no control over which sort of squadron they were assigned to, these instruments represent exogenous treatment statuses that impacted achievement.

Columns 2, 3 and 4 of Table 1 show student summary statistics across the three types of squadrons. Around 50% of students were in the control squadrons. They are not statistically different from the overall population on any background measure. 35% of students are in bifurcated (Treatment-B) squadrons. These squadrons are largely made up of students with high verbal aptitude or students in the bottom tercile. The remaining 15% of students are in homogeneous (Treatment-H) squadrons. These squadrons are comprised of the other students who were neither high verbal ability nor bottom tercile. For all academic measures, the standard deviation among students is lowest within homogeneous squadrons. The three squadron types are balanced along racial and gender demographics. Even in to sophomore year (the treatment status was only for freshman year), students from the homogeneous squadrons have a higher mean GPA than those from control and bifurcated squadrons.

4.3 Social Network Survey and Patterns of Friendship Formation

In the spring semester of 2010, when the classes of 2011 and 2012 were juniors and sophomores, respectively, students were administered a survey asking them to list up to five "friends" and five "study partners" from their sophomore years.[20] Friend and study partner nominations did not need to be mutually exclusive or reciprocal.

There was no requirement that friends and study partners named on the survey needed to be in the same squadron, but this was overwhelmingly the case. Over 79% of all nominations were intra-squadron, lending credibility to the idea that squadrons at USAFA form discrete peer groups for students. Athletes were some of the most likely students to list connections from other squadrons. These nominations will form the basis of 80 distinct social networks, 40 squadrons across 2 cohorts.

549 students completed the survey for a response rate of 25%.[21] I impute the full social networks using a modified version of Chandrasekhar & Jackson (2012)'s model of social network reconstruction. The full imputation process is detailed in Appendix A, but the underlying assumption is that the observed social network information will be representative of the overall network. Even with a 25% response rate, I am able to effectively reconstruct the complete social networks of the squadrons. The ultimate goal is to have consistent and precise estimates of contemporaneous and contextual peer effects. In Appendix B, I estimate contemporaneous and contextual peer effects using a separate data set where the whole social network is observed. I then simulate a 25% response rate, perform the imputation and show that resulting estimates of contemporaneous and

[20] The exact wording of the survey read *Please enter the first and last names of up to five cadets you most frequently studied with as a third degree cadet.* and *Please enter the first and last names of up to five cadets with whom you spent most of your free time during your third degree year.* Third degree refers to a students sophomore year at USAFA.

[21] Students at USAFA have a graduate academic adviser, akin to a teaching assistant, who is randomly assigned to students within majors. Students all meet with their adviser at the same time towards the end of each term. The survey was administered by these advisers to their group of students during the Spring 2010 meeting. Certain graduate academic advisors neglected to have their students sign the consent form, these completed surveys had to be discarded. Other advisers failed to administer the survey at all. There is little reason to believe that the neglect of graduate advisors is correlated with student characteristics. Selection into survey taking is controlled for using an inverse propensity score weight nonetheless.

contextual peer effects using the imputed sample are not significantly different from the estimates using the full observed network.

Column 5 of Table 1 shows the summary statistics for those who responded to the survey. Overall, they are a positively selected group in terms of academic proficiency. Table 2 shows average responses from the survey for either sort of connection and for study partners and friends separately. Respondents named an average of 2.7 peers in their squadron as connections. The reciprocity rate among connections was 46% overall, 55% for study partners and 59% for friends. In total, 67% of all students were named on at least one survey.

Table Table 3 shows patterns of connection formation among student types. Each column reports a separate logistic regression. For survey responder, there is one observation for each other member of their squadron, with the dependent variable indicating whether or not they were listed as a connection.[22] Independent variables represent information about the pair of students (e.g. both female, one athlete and one non-athlete). Results account for the frequency of different student types at USAFA and tell a store about the relative propensity of connection formation between two students.

In Column 1, friendships and study partners are treated equally. We see that a connection between two females is significantly more likely to occur than between two males (the omitted category) and co-ed connections are less likely. Athletes are less likely to be connected with non-athletes. Race is not a significant predictor of a connection. The bottom rows describe the likelihood of connections among students from the different terciles of the predicted ability distribution (a connection between to top-tercile students is the omitted category). Top students seem to be the most insular, with connections between bottom and top tercile and medium and top tercile being significantly less likely to occur.

When comparing differences in the likelihood of study partnerships versus friendships forming,

[22]Equation 8 in the appendix formally defines the model producing Table 3. It is discussed here for descriptive purposes about friendship formation, but coefficients from the models are also used in the network imputation process to provide probabilities for different sorts of connections occuring.

an interesting pattern emerges. Study partners, it seems, end up being formed less based on typical demographics and more based on academic ability. Women and men are more likely to study together than to be friends. Study partners are more regularly interracial and marginally more likely to involve an athlete and a non-athlete. Looking at predicted ability, a friendship between a predicted bottom tercile and predicted top tercile student is not significantly less likely than one between two top-tercile ones. However, study partnerships are significantly less likely to form among students from these two groups. These patterns are not causal, but speak to the sort of social cliques (e.g. a clique among top students) that can result in spillovers in attainment being concentrated among small, tightly connected group of students while other, less well connected students cannot benefit.

5 Results

The experimental sorting of students during their freshman year provides the exogenous variation used to identify contemporaneous and contextual effects in sophomore year. Results from the sorting are presented in Table 4. In freshman year, students were either in control (omitted category), bifurcated or homogenous squadrons. The bifurcated squadrons had non-monotonic effects on students, harming predicted low ability students and aiding predicted high ability ones. For this reason, I interact whether a student was predicted low or high ability with the indicator for being in a bifurcated squadron. The impacts represent exogenous shifts in student attainment due to experimental sorting of students. In predicting a student's sophomore year GPA, I control for own characteristics, peer characteristics as well as an indicators for predicted tercile and an indicator for being a high-achieving verbal student, one criteria on which the sorting was based.

The experiment is a significant predictor of achievement in sophomore year, with students from homogeneous squadrons predicted to increase their sophomore GPA by 0.075. Top tercile students in the bifurcated squadrons performed better, relative to their counterparts in the control squadron by 0.069 GPA points and bottom tercile students were harmed, experiencing a -0.061

GPA points penalty of the bifurcated squadrons. The standard deviation of sophomore GPA is .547 so these shifts are over a tenth of a standard deviation in magnitude. The joint F-statistic on the experimental indicators is 7.46. The probability of falsely rejecting the null that all indicators are zero is 0.0002.[23] The coefficient estimates from Table 4 are used to generate predicted attainment in sophomore year for all students, $\widehat{Y}_{ic}$.

Using results from the first stage, I move to estimating social effects among students in their sophomore year. Table 5 builds up to my preferred specification. To start, two students are considered connected if they are either study partners or friends. All specifications control for individual background characteristics. In Column 1, I estimate contemporaneous effects using *actual* mean friend achievement, GY_{ic} and OLS. Referring to the coefficient as a contemporaneous effect is slightly misleading since it is not identified. The estimate encompasses the combined effect of peer spillovers along with selection into friendships. The contemporaneous effect coefficient is 0.14. If connections among USAFA students are, on average, reinforcing differences in ability (i.e. above average students tending to connect with other above average ones and same for below average), we would expect this unadjusted coefficient to be larger than the instrumented one. As can be seen in Column 2, this is, in fact, the case. Switching from unadjusted GY_{ic} to the instrumented $\widehat{Y}_{ic}$ lowers the magnitude of the contemporaneous effect by around half. This specification does not control for contextual contextual effects, and so is still not identified. The reduction in magnitude between Columns 1 and 2 shows that the instrument shifts the estimate of contemporaneous effects, consistent with expectations.

Interpretation of significance requires extra care going forward to account for imputation. Most presented results, such as those in Table 5, involve imputed social connection data. These are results obtained from running the same model multiple times, once for each imputed social network.[24] Each reported estimate contains three values. The first is the average point-estimate. The

[23] Stock et al. (2002) show that a first-stage with three instruments would need a joint F-statistic of 9.08 to ensure bias is no more than 10% of OLS. My F-statistic would reduce that confidence threshold to around 15%

[24] Appendix A details the steps of network reconstruction and Appendix B tests the validity of my imputation methodology.

second is the average standard error, capturing within model variance. Significance stars treat the average point-estimate and average standard error as if they were from a single regression. The third is the standard deviation of the point-estimates. This captures the variance in my estimates across models.

For hypothesis tests from models that involve multiple repetitions, I follow Schafer (1997)'s formula for calculating p-values in the presence of data augmentation.[25] Briefly, this proportionately weights both the within and across repetition variation and calculates total variance. Total variance will always be larger than either the average standard error or standard deviation of estimates. It provides a conservative estimate of statistical significance, which is appropriate given the imputation, and will be the default for discussion of hypothesis tests.

Column 3 of Table 5 estimates Equation 2 and presents the relationship between peer background characteristics and performance. As discussed in Section 3, without separate identification of the contemporaneous channel of peer effects, these results represent the combined contemporaneous and contextual channels. They do not identify mechanisms underlying the peer effects. Instead, the give a picture of whether certain types of social connections (study partnerships or friendships) are associated with stronger performance. A few results are statistically significant, but none are of a sizable magnitude. For example, the academic entry score of an individual's social ties is negatively correlated with own attainment. Adjusting for scale, a student switching social ties to a group with a one-standard deviation higher average academic entry score would cause a predicted drop in own attainment of .02 GPA points.[26] Two other significant coefficients are percentage of social ties attending a military preparatory academy and percentage black, but the magnitudes are similarly small. Preparatory school and black have population means of .17 and .05 respectively and 55% of black students at USAFA attended a preparatory academy so there is

[25]They define the total variance of the model as $T = \bar{U}+(1+m^{-1})B$. Where m is the number of repetitions, $\bar{U}$ is the average standard error and B is a measure of the between-imputation variance. Inference is based on $T^{-\frac{1}{2}}(\beta_0-\bar{\beta}) \sim t_v$

[26]Academic entry is normalized to mean 0, standard deviation 1 at the student level. The standard deviation across squadrons is 0.21. The point estimate of -0.11 represents the predicted change in attainment if peer academic entry score were to change by 1.

some overlap in the predicted effects.

Column 4 estimates both contemporaneous and contextual effects. This is the full estimation of Equation 6 and is my preferred specification. The contemporaneous effect, representing spillovers in attainment driven through a student's social connections, are positive and significant.[27] It is estimated at 0.108, which implies a social multiplier from connections effort of 1.12. For context, a student who receives an exogenous shock to ability (e.g. outside tutoring) that raises own attainment by 1 GPA will create 0.12 GPA points worth of spillovers among his peer group. The spillovers are predicted to be heavily concentrated among a student's closest friends, quickly decaying as you move further away in the social network.

The full list of estimates from Column 4 can be seen in Table 6. Contextual effects are either insignificant or of relatively small magnitudes. Bramoulle, Djebbari & Fortin (2009) and Lin (2010) also find that the contemporaneous channel of peer effects is the driving force and dominates contextual ones. Students seem much more responsive to the efforts of their connections than to their background characteristics or pre-USAFA ability.

Next, I separately estimate peer effects by type of social connection. Column 1 of Table 7 repeats the main specification while Columns 2 and 3 show results for study partners and friends separately.[28] Focusing on the contemporaneous effect, study partners exhibit spillover rates of 0.133 versus 0.043 for friendships. For study partners, using Schafer (1997)'s formula for total variance, the estimate is significant at the 15% level. What is clear, is that study partners are more prolific in their creation of spillovers onto other study partners.

Study partners, imply a connection between peers that is somewhat focused on academics so the result is intuitive. It is also exciting because study partners represent a far more specific sort of relationship than does friendship. These students are sitting down together and collaborating.

[27] A p-value test accounting for total variance gives a p-value of 0.09.

[28] The imputations of missing connections were also done separately. In short, to perform the study partner imputation, no information on friendship nominations were used and visa versa. Each of the 50 repetitions of network impuation for study partnerships is independent from the 50 repetitions based on friendships. Appendix A has full details.

Peer effects are largely contemporaneous, which implies that studying benefits both parties, not just from higher ability to students to lower ability ones. Students studying and working together seems a necessary condition for spillovers, versus the alternative theory that students may react to role-model students in their peer group from afar. Assuming that friendships are directed (i.e. non-reciprocal), shown in Column 4, does not alter estimates in a significant way.

My estimates of the size of the contemporaneous effect, 0.108, are smaller than similar estimates from Lin (2010) (0.274) and Bramoulle, Djebbari & Fortin (2009) (0.467).[29]. Their papers study representative population of high school students. There are two main differences in my setting from theirs that likely cause that differing estimates of contemporaneous effects. First, the students at USAFA are not college students, not high schoolers and are nationally representative. Second, I am able to instrument for average GPA of connections using experimental variation rather than relying on the strong spacial assumptions that come with using friend-of-friend performance as an instrument for friends.

I cannot conclusively show why my results differ, but Table 8 presents some evidence. I switch from using my set of instruments derived from the experimental sorting and instead use a student's connections-of-connections, the same as used by prior authors, as an instrument for the mean performance of their connections. The columns in Table 8 correspond with those in Table 7. All estimates remain insignificant so no strong conclusions can be drawn, but the point estimate of the endogenous effect does nearly triple to 0.299. This is suggestive, but not conclusive, that the friend-of-friend instrument may actually exaggerate the endogeneity of friend selection rather than mitigate it.

So if social connections facilitate spillovers among peers, who stands to benefit the most? Unlike in the linear-in-means model, not all students equally gain from being placed in an ideal peer group. Rather, it is the students who make the strong social connections who stand to benefit. To ensure that the highest percentage of students possible form beneficial social ties, it would make

[29] Bramoulle, Djebbari & Fortin (2009) is estimating peer effects on participation in extracurricular activities, rather than GPA

sense to try to maximize the overall *potential* connections in a squadron.

I examine whether or not being randomly assigned into a squadron with more potential friends causes a student to do better or worse. The results are suggestive, not causal. I use the patterns of friendship formation from Column 1 of Table 3 to estimate the probability of every pair of students from the same squadron forming a tie. I sum these probabilities for each student to estimate a predicted number of social ties within their squadron and adjust for relative frequencies of students by gender and race.[30] Predicted connections are a function of random squadron assignment and do not correlate with a student's characteristics or ability. Table 9 replicates my preferred specification by connection type, but weights students based on this measure of predicted social ties. Estimates of the contemporaneous effect rise by 8-10% for the combined, study partner and directed specifications. None of the estimates are statistically different from their unweighted ones, but it does suggest that contemporaneous spillover effects are larger among students with more predicted friends. The larger estimates of contemporaneous effects could be indicative of students with more predicted friends tending to have more or stronger actual social ties and thus experiencing larger rates of spillovers. This would be in line with Patacchini et al. (2012), who have a measure of the strength of social ties and find that effects from strong ties among students dominate those from weaker ties.

6 Conclusion

Understanding the mechanisms that cause peer effects informs policy makers and administrators interested in maximizing spillovers among students. This paper advances our knowledge of how spillovers manifest, by showing that close social connections, specifically study partners, create spillovers in achievement. My findings underscore the importance of social integration for students. A student will not experience beneficial spillovers from their peers unless they are able

[30]For each gender and racial group, I subtract the average number of predicted friends among all students in that group from each student's own prediction. This adjusts for the fact that say, white males are always going to be predicted to have a higher number of friends due to patterns of homophily.

to make meaningful social ties.

To date, the bulk of peer effects research has used a version of the linear-in-means model.[31] It identifies peer effects by comparing outcomes across multiple groups of students. This model is limited because it cannot separate the two major channels of peer effects: contemporaneous and contextual effects. This limitation has forced researchers to focus on whether or not peer effects exist, rather than on their causes.

In my paper, I use a social network model similar to Bramoulle, Djebbari & Fortin (2009) and Lin (2010) to estimate peer effects. The model generalizes the linear-in-means model and assumes that students are only directly affected by their friends and study partners, rather than by all of their peers. Under this approach, there is substantial within-group variation in students' social connections that can separate contemporaneous effects from contextual ones.

My approach builds on prior social network models. I use experimental variation in student grouping in a pre-period as an instrument for achievement. Prior studies of peer effects using social networks have typically only had cross-sectional data, requiring them to rely on stronger assumptions with regards to their exclusion restrictions. Second, my data contain student nominations on friendships as well as study partners. This is the first time, to my knowledge, that study partnerships have been used in estimating a social network model of peer effects. This is particularly important when thinking about mechanisms. The contemporaneous channel of peer effects can encompass many possible mechanisms, but the nature of my data narrows in on one.

Results are significant for several reasons. I find a contemporaneous effects tend to dominate contextual ones. I estimate a contemporaneous effect of 0.1, which implies a social multiplier of attainment of 1.1. Study partners, rather than friends, drive the effect. My results are the first causal estimates of contemporaneous peer effects on attainment in higher education using social network data. Evidence of a contemporaneous effect alone is important. It shows that students respond to the efforts and achievement of their close social ties, not just to their environment.

[31] As mentioned previously, the use of the linear-in-means model has been data driven. Data on how students are grouped is relatively available, while data on social connections within groups, relatively sparse.

The contemporaneous effect is driven by the social network of study partners. It is insignificant when estimating a social network based only on friendships. The strength of the effect from study partners offers a concrete mechanism for how and when spillovers manifest among students. Study partners are a specific kind of relationship. It indicates that students spend at least some amount of time working together on academics.

Evidence from linear-in-means studies, which far outnumber social network ones, suggests that higher-ability peers lead to better outcomes. My results provide a reason behind this. It is not that high-ability students will automatically rub-off on their peers and generate positive spillovers. Students who are grouped with higher-ability peers perform better on average because they are more likely to make a beneficial social connection. This supports findings from Fletcher et al. (2013) and Booij et al. (2015), which do not spillovers, but do find evidence that students perform better when placed in classes with a larger supply of potential friends.

The takeaways from my findings speak to how difficult social engineering is from a policy or administrator perspective. While arranging students into groups is relatively easy, creating social ties among them is not. There are patterns in social tie formation that can be leveraged, which are guided by student's propensity to form social ties with similar students. Gender and race are two large explainers of social ties, but I show that athletic and academic ability can play a role as well. I provide suggestive evidence that students who are randomly placed into a peer group with a higher number of expected study partners (calculated based on observed patterns of study partner formation) experience larger contemporaneous effects on attainment. The key is not that someone's study partner is smarter than them, the key is that students are studying together at all. For students with a higher chance of finding a study partner, the spillover effects are larger. Arranging students into groups in such a way as to encourage collaboration, but not disrupt the social fabric of a school, is a subject of future research.

This paper isolates two mechanisms that contribute to peer spillovers, but we are still a long way from understanding the totality of what happens within groups of students to create spillovers.

In particular, the question of *when* these spillovers occur requires more detailed information on how the social ties among students form and break over time. A cross-sectional snapshot of the social network, like the one studied in this paper, cannot account for the dynamics of educational social environments. Understanding such intricacies will be necessary for peer effects research to provide reliable and replicable results for policy makers and administrators.

Appendices

A Network Imputation and Simulation Results

This section details the steps taken to simulate the complete social USAFA social network in the presence of survey information from only a subset of students. The methodology is adopted from Chandrasekhar & Jackson (2012), but modified to account for non-random selection into survey taking. The steps can be summarized as:

1. Estimate selection into survey taking, create an inverse propensity score weight.
2. Using that weight and dyadic data set, estimate the probability of friendships formation across student types
3. Estimate weighted triad probabilities across student types
4. Following Chandrasekhar & Jackson (2012)'s method of network reconstruction for sparse networks, use estimates from step 2 to simulate a complete USAFA squadron 250 times.
5. Estimate Equation 5 for each repetition from step 3 and report results as distribution of parameters, shown in Section 5.

1. Selection into Survey Taking: The 25% of students who completed a friendship survey were not a randomly selected group of individuals. As shown in section 4, survey takers were, on average more likely to be high achievers, white, male and non-athletes. Chandrasekhar & Jackson (2012) assume knowledge of a random subset of individuals in reconstructing social networks, which with my data would lead towards biased probability estimates of friendships formations. To adjust, I estimate the following probit model for survey taking:

$$P(Taker_i = 1) = \Phi(\beta_0 + \beta_1 X_i) \tag{7}$$

Where $Taker_i$ is an indicator for whether or not student i filled out a survey and $P(Taker_i = 1)$ is the underlying probability. X_i is a vector of background characteristics that includes sex, race, sat scores, academic and leadership composites and cohort. Results are shown in Table A1. As evidence by the differences in summary statistics, survey responders were positively selected academically (higher SAT Math and academic composites), less likely to be athletes and not significantly different with regards to race, gender, or experimental squadron type. The average $P(Taker_i = 1)$ is 0.25 and ranges from 0.046 to 0.491.

2. Estimating Friendship Probabilities: Not all friendships are created equal, or equally likely. It is well documented[32] that friendship formation tends to exhibit homophily - students forming friendships with other similar students. Typically, gender and race are two of the largest drivers of homophily. So when modeling friendship formation it is important to capture student's tendency towards homophily (as well as tendencies to for cliques, discussed later) by assigning different probabilities to varying sorts of friendships.

To calculate friendship probabilities I run a regression of survey takers' possible dyads. The data are setup in a dyadic fashion with a single observation represented as $Connected_{ijc}$. Students are either "takers" (i.e. they completed the friendship survey) or non-takers. D_{ijc} contains the set of dyads with $i \in takers$ and $j \in takers, non\text{-}takers, \neq i$. If a particular squadron has 26 students and 10 survey takers, there would be 250 dyad observations for the squadron, with each survey taker matched with the 25 other students in the squadron. $Connected_{ijc} = 1$ if student i listed student j as a connection and 0 otherwise. Characteristics of the pair of students D_{ij} are all binary and describe the pairing (i.e. same-gender, same-race, both survey takers, both-athletes, etc...). Relative frequencies of friendship formation at USAFA are then estimated with a probit model weighted by the inverse-propensity weight from the previous step.

$$P(Connected_{ijc} = 1) = \Phi(\lambda_0 + \lambda_1 D_{ijc}) \quad (8)$$

[32] See Jackson et al. (2009), McPherson et al. (2001) among others

Table 3 shows estimates from this model. Friendship patterns among survey responders at USAFA follow some typical patterns when it comes to gender, but not race. Gender is the single strongest predictor of friendship. Male-Male friendships are the omitted category and we can see that Female-Female friendships are significantly more likely to occur, which makes sense given that females are in the minority. Female-Male friendships, on the other hand, are far less likely to occur than Male-Male. These findings will be particularly useful in simulation of the complete network as we can be confident that, more often then not, the handful of females within a squadron will be friends with relatively few co-ed friendships linking them to the males in their squadron. Athletes are not significantly different making friends with each other compared to non-athletes, but they are less likely to be friends with a non-athlete. When it comes to race, there is no significant difference in the propensity for two minority students, two white students or one minority and one white student to choose each others as friends.[33] This stands as a notable contrasts to homophily studies on the nationally representative AddHealth data set such as Bramoulle, Rogers, Currarini, Jackson & Pin (2009), where race is the primary factor in student homophily.

Coefficients on incoming predicted ability showcase what was at the heart of Carrell et al. (2013) counterintuitive findings. Estimates are relative to two top-tercile students being nominated as friends. Top-tercile students are significantly less likely to be friends with students from the middle or bottom-tercile of predicted ability. Bottom and middle-tercile students show no significant differences in friendship formation relative to two top-tercile students. Lastly, coming from the same freshman squadron is a strong predictor of sophomore friendship. This is disconcerting from an identification perspective as it means the freshman year experiment instrument would violate the exclusion restriction for such friends. However, the average student has only .66 students from their freshman squadron in their sophomore one, less than 2% of potential sophomore friendships. Dropping these friendships does not impact results.

[33] Alternate specification looked at all minority races separately (e.g. Hispanic-Asian, Black-White, Black-Black, etc...), not reported hear. The larger number of possibilities led to less precision. Relative to a friendship between two white students, white-black friendships were significantly less likely and Black-Hispanic friendships significantly more. No other racial groups showed significant differences in friendship formation likelihood.

3. Estimating Friendship Cliques: Key to reconstructing a realistic social network is the fact that links are not formed independently. From both common sense and empirical observation (See Jackson (2010)), it is found that people tend to form cliques. Observationally, if student a is friends with both students b and c, then b and c will be friends a much higher percentage of the time than if friendships were formed independently. To capture social cliques, Chandrasekhar & Jackson (2012) recommend using subgraphs. A subgraph is any pattern of links that can be formed with a subset of the nodes in a network. The authors show that the frequency of triads (3 people who are all linked) is very effective at capturing social clique phenomena.[34] So while Equation 8 captures the likelihood of any two students being friends, based on observables, we also need to incorporate separate probabilities for complete triads that are different from the product of three independent friendships. To do this, data are arranged in triadic form T_{ijkc} where all $i, j, k \in takers$ and $T_{ijkc} = 1$ if all three are linked.

Triads are grouped into seven mutually exclusive categories, each with a row in Table A2.[35] The groups were chosen to be balanced in terms of potential triads and also to depend on gender, ability and race, three factors important in choice of friends. In total, there are 235,074 potential triads among students in the same sophomore squadrons. Column 1 shows the percentage of triads corresponding with each group. Column 2 shows how often the complete triads occur within the survey data, conditional on triads with at least two of the students responding to the survey and weighted by the survey responders' inverse propensity scores. I call this P_{gijk}, the probability of a complete triad occurring for students i, j and k who belong to group type g. Column 3 shows how common each complete triad type is, relative to a baseline frequency if links were

[34] The ratio at which complete triads form above the expected rate if friendships were independent can almost be thought of as a sufficient statistic for the "clique-ishness" of the network. Of course, there are other patterns of friendship formation that could be considered (four people who are all friends form an "X-box" and five would form a pentagram when laid out visually), but estimating their likelihood of formation conditional on complete triad rates has not been found to contribute much to the accuracy of reconstruction models while posing considerable computing burdens.

[35] The groups are: All female (the scarcity of females makes it intractable to separate their triads by race and ability), mixed gender (either one or two males) with a "High-Low" friendship (i.e. a predicted top-tercile and predicted bottom-tercile student), mixed gender without high-low, all male and all white with high-low, all male and all white without high-low, all male mixed race with high-low and all male mixed race without high-low.

generated independently. Among survey takers, 0.13 of possible dyads were listed as friends. If friendships were independent of each other, the baseline likelihood of observing a complete triad would be $0.13^3 = 0.0023$. Column 3 shows Column 2 divided by 0.0023 and shows how much or (less) common complete triads occur relative to baseline. Unsurprisingly, complete triads among females occur 20 times more regularly than baseline. There is not a large quantity of all female triads possible in the data, but those that do occur are likely to see all three females list each other as friends. Mixed gender complete triads occur slightly less or slightly more often than baseline, depending on whether there is or is not a friendship between a top-tercile and bottom-tercile student. All male friendships, measured across multiple racial and ability categories always occur more frequently than the baseline would predict. All white complete triads with no top-tercile and bottom-tercile student are the most regular among these.

4. Create Reconstructed Networks: At this point, I have P_{ij} and P_{gijk} which give the probability of a single link forming among two students and the probability of a complete triad (three links) forming among three students, respectively. To reconstruct the networks I follow Chandrasekhar & Jackson (2012) model for sparse networks. This involves first choosing an overall density of the network. I obtain this by assuming that survey responders are representative for the *number* of unique friends listed on their survey. For undirected friendships, I assumed the average student would have 3.1 friends. Next, I specify how many complete triads (three students who are all friends), N_t, and how many unsupported friendships (friendships that are not a part of a complete triad), N_f, the average squadron should have. N_t and N_f were chosen so that the total number of friendships would give the desired density and so that the ratio of triads to unsupported friendships would conform to the relative frequency of triads shown in step 3.

The reconstruction process, which I repeat 250 times is as follows: Starting with only the known friendships given by survey responders, friendships are added one at a time based on P_{ij}. After each friendship is calculated, a count is taken of the number of unsupported friendships, n_f and the number of complete triads n_t. It is possible that adding a single friendship may create a complete triad and potentially turn previously unsupported friendships into triads so the counts do not necessarily increment by 1 each addition. Once N_f is reached, complete triads are added, drawn based on P_{gijk} and added. After each addition, n_t and n_f are recalculated. Adding a triad will increase n_t, but could reduce n_f. If n_f dips below N_f, individual friendships are added until N_f is again reached. This continues until N_t is reached at which point the draw of the social network is complete. This is done separately for the two cohorts of data, but simultaneously for all squadrons of the same cohort.

5. Estimate Model of spillover and Contextual Effects: Each time step 4 is completed, there is a new draw of G^m, the social network at USAFA. Each of these are used to estimate the models discussed in the Methods Section.

B Simulations on the Add Health Data Set

To the best of my knowledge, the empirical section of Chandrasekhar & Jackson (2012) is the only application of their reconstruction method to date. While it shows the promise of the method, there are sufficient differences between our settings that it should not be taken for granted that their model will yield unbiased results in my setting. The first difference is the selection into survey taking. Chandrasekhar & Jackson (2012) assume that the friendship nominations surveys were a random subset of their population. I know for a fact that the USAFA surveys were not taken at random and so account for this using an inverse propensity score weight described earlier. The second difference in our settings is size. While I estimate a model for 80 social networks of around 30 students each, they estimate their model on approximately 40 Indian villages with an average size of 2,000 households.

My approach is to use The National Longitudinal Study of Adolescent Health (AddHealth) where the “population” values of a group of social networks is known in order to test the effectiveness of network reconstruction in my setting. AddHealth is a nationally representative school-based, longitudinal study of the health-related behaviors of adolescents and their outcomes in young adulthood. The study's sampling design was to target full populations within schools. The study includes 80 high schools and 54 feeder schools. The median cohort within a school consists of 275 students. An in-school survey was given to students that asked them to nominate up to five male and five female friends.[36]

It is not possible to estimate my preferred specification from Equation 5 since the AddHealth data does not have randomized assignments like USAFA. Instead, I perform reconstruction using Lin (2010)'s model for estimating the endogenous peer effect using social networks. The model is close to my own and should be a viable proxy. Since size is one of the primary differences between USAFA and Chandrasekhar & Jackson (2012), I create a subset of the AddHealth data that looks more like USAFA. I start with the 67,000 observations used by Lin (2010) and use the same assumption of cohorts within schools forming distinct social networks. I then limit the data to cohorts of between 19 and 69 students and to schools where the response of completing the survey was 85% or better. This results in 40 cohorts and a subpopulation sample of 1,813 students.

Using this subpopulation of AddHealth data, I simulated the partial-response environment of USAFA. First, I assigned each AddHealth student a simulated probability of completing the friendship survey based on point estimates from Equation 7 run on the USAFA sample. 456 (25%) students were selected using a weighted random draw as simulated survey takers. The friendship nomination data for the other 75% of students was discarded. Using the 25% of friendship nominations, I performed each step of the network construction from Appendix A. This included calculating selection into survey taking, estimating the probability of friendship and triad formation and then running the Chandrasekhar & Jackson (2012) model for network reconstruction for 50 independent repetitions. Each of the 50 reconstructed social networks was then used to separately estimate Lin (2010)'s social network model.

Table B1 shows the results from this simulation. Column 1 contains estimates from the reconstructed sample. Each variable has three values associated with it. On top is the average point

[36] See Lin (2010) and Patacchini et al. (2011) among others for full descriptions of the data set.

estimate, measuring the relationship to a student's GPA. These can be interpreted in the same fashion as point estimates in my main results. Below, in square brackets, is the average standard error across the 50 repetitions. Below that, in parentheses, is the standard deviation of the point estimate across the repetitions. The point estimates themselves are not of much interest, rather it is their relationship to Column 2. Column 2 presents the output of running the social network model using all of the available (approx 95% response rate) friendship nominations for the subpopulation of AddHealth students. This can be thought of as the "population" values that Column 1 is attempting to reconstruct. Overall, the reconstruction performs well. The top row, Friend GPA is the most important coefficient, representing the endogenous peer effect. The AddHealth population value is 0.318 and the average values among the reconstructed networks is 0.305 with an average standard error of 0.039 and standard deviation of 0.046, meaning the true value is contained within even the most conservative of standard confidence intervals. Own characteristics were assumed to be known for all observations and so predictably do not deviate much at all from their population values. Friendship background characteristics, analogous to exogenous effects, are largely insignificant among the reconstructed and subpopulation samples, but signs always point in the same direction. Column 3 is a replication of Lin (2010)'s model 6, using the full Addhealth population.[37] Figure A1 graphically conveys the distribution of estimates using friendship imputation relative to estimates using the full population of Add Health friendship nominations.

Using Schafer (1997)'s formula for inference for multidimensional models with imputation, I calculate a more formal test of the validity of the model. The formula incorporates both sources of variation present in the reconstructed models - the variation across estimates and the variance matrices within each model. I will not recreate the formulate here, but the end result is the ability to obtain p-values to test the null hypothesis that the estimated values from Column 1 are the same as those in Column 2. I calculate p-values for three sets of regressors: only the endogenous effect; the endogenous and all exogenous friendship effects; lastly all estimated coefficients.

[37] I want to note that the author making her code available was a huge help, saving me much time and providing an easy path for replication

The respective p-values for the null hypothesis are: .65, .71, and .97. In each instance I fail to reject the null. The primary concern of these replication was that the smaller networks of USAFA, relative to Chandrasekhar & Jackson (2012)'s empirical sample and the non-random selection into survey taking may have introduced biases in the reconstruction process. The large p-values, calculated to account for multiple imputation are a strong sign that, even in the face of 25% of a network sample, utilizing the underlying patterns in friendship formation, social network structure can yield unbiased estimates regressors to be included in peer effects models.

References

Bifulco, R., Fletcher, J. M. & Ross, S. L. (2011), 'The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health', *American Economic Journal: Economic Policy* **3**(1), 25–53.

Booij, A. S., Leuven, E. & Oosterbeek, H. (2015), Ability Peer Effects in University: Evidence from a Randomized Experiment, SSRN Scholarly Paper ID 2554917, Social Science Research Network, Rochester, NY.

Bramoulle, Y., Djebbari, H. & Fortin, B. (2009), 'Identification of peer effects through social networks', *Journal of Econometrics* **150**(1), 41–55.

Bramoulle, Y., Rogers, B., Currarini, S., Jackson, M. O. & Pin, P. (2009), Homophily and long-run integration in social networks, Discussion Paper 1475, Northwestern University, Center for Mathematical Studies in Economics and Management Science.

Calvó-Armengol, A., Patacchini, E. & Zenou, Y. (2009), 'Peer effects and social networks in education', *The Review of Economic Studies* **76**(4), 1239–1267.

Carrell, S. E., Fullerton, R. L. & West, J. E. (2009), 'Does your cohort matter? measuring peer effects in college achievement', *Journal of Labor Economics* **27**(3), 439–464.

Carrell, S. E., Page, M. E. & West, J. E. (2010), 'Sex and Science: How Professor Gender Perpetuates the Gender Gap', *The Quarterly Journal of Economics* **125**(3), 1101–1144.

Carrell, S. E., Sacerdote, B. I. & West, J. E. (2013), 'From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation', *Econometrica* **81**(3), 855–882.

Chandrasekhar, A. & Jackson, M. O. (2012), Tractable and consistent random graph models, SSRN Scholarly Paper ID 2150428, Social Science Research Network, Rochester, NY.

Chandrasekhar, A. & Lewis, R. (2011), 'Econometrics of sampled networks'.

Christakis, N. A., Fowler, J. H., Imbens, G. W. & Kalyanaraman, K. (2010), An empirical model for strategic network formation, Working Paper 16039, National Bureau of Economic Research.

De Giorgi, G., Pellizzari, M. & Redaelli, S. (2010), 'Identification of social interactions through partially overlapping peer groups', *American Economic Journal: Applied Economics* **2**(2), 241–275.

Duncan, G. J., Harris, K. M. & Boisjoly, J. (2001), 'Sibling, Peer, Neighbor, and Schoolmate Correlations as Indicators of the Importance of Context for Adolescent Development*', *Demography* **38**(3), 437–447.

Feld, J. & Zoelitz, U. (2015), 'Gender peer effects in higher education'.

Fletcher, J. M., Ross, S. L. & Zhang, Y. (2013), The determinants and consequences of friendship composition, Working Paper 19215, National Bureau of Economic Research.

Foster, G. (2006), 'It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism', *Journal of Public Economics* **90**, 1455–1475.

Haynie, D. (2001), 'Delinquent Peers Revisited: Does Network Structure Matter?', *American Journal of Sociology* **106**(4), 1013–1057.

Hoxby, C. (2000), Peer Effects in the Classroom: Learning from Gender and Race Variation, Working Paper 7867, National Bureau of Economic Research.

Jackson, M. O. (2010), *Social and Economic Networks*, Princeton University Press.

Jackson, M. O., Currarini, S. & Pin, P. (2009), 'An economic model of friendship: Homophily minorities and segregation', *Econometrica* .

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

Liu, X., Patacchini, E., Zenou, Y. & Lee, L.-F. (2012), Criminal Networks: Who is the Key Player?, Working Paper 2012.39, Fondazione Eni Enrico Mattei.

Manski, C. F. (1993), ‘Identification of endogenous social effects: The reflection problem’, *The Review of Economic Studies* **60**(3), 531–542.

Marmaros, D. & Sacerdote, B. (2006), ‘How Do Friendships Form?’, *The Quarterly Journal of Economics* **121**(1), 79–119.

McPherson, M., Smith-Lovin, L. & Cook, J. M. (2001), ‘Birds of a Feather: Homophily in Social Networks’, *Annual Review of Sociology* **27**, 415–444. ArticleType: research-article / Full publication date: 2001 / Copyright 2001 Annual Reviews.

Patacchini, E., Rainone, E. & Zenou, Y. (2011), Dynamic aspects of teenage friendships and educational attainment, CEPR Discussion Paper 8223, C.E.P.R. Discussion Papers.

Patacchini, E., Rainone, E. & Zenou, Y. (2012), Student networks and long-run educational outcomes: The strength of strong ties, CEPR Discussion Paper 9149, C.E.P.R. Discussion Papers.

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

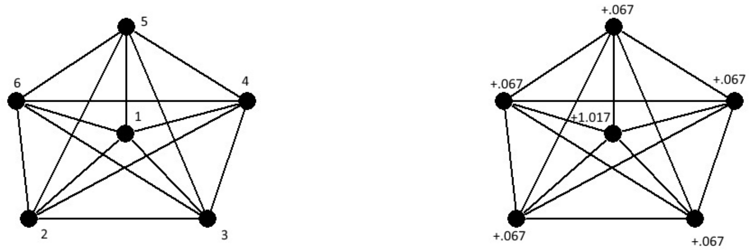
Schafer, J. (1997), *Analysis of Incomplete Multivariate Data*, 1 edition edn, Chapman and Hall/CRC.

Stock, J. H., Wright, J. H. & Yogo, M. (2002), ‘A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments’, *Journal of Business & Economic Statistics* **20**(4), 518–529.

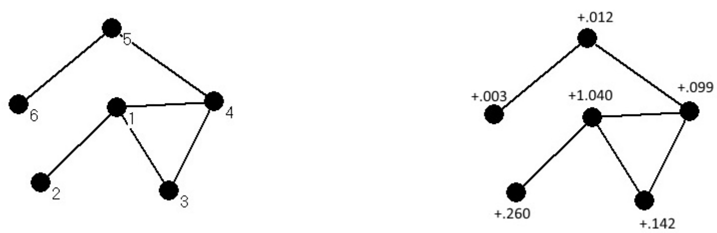
Zax, J. S. & Reese, D. I. (2002), 'IQ, academic performance, environment, and earnings', *The Review of Economics and Statistics* **84**(4), 600–616. ArticleType: research-article / Full publication date: Nov., 2002 / Copyright 2002 The MIT Press.

Tables and Figures

Figure 1: Spillover Examples



(a) Spillovers With Peers Assumption



(b) Spillovers With Social Network Assumption

Left picture of panel (a) and (b) show the connections among students (labeled 1-6) under the assumptions of a social network and peer effects model respectively. The right hand figures show the resulting predicted increase in GPA caused by a one-unit exogenous increase in student 1's performance

Table 1: Student Summary Statistics

	(1) All	(2) Control	(3) Treatment-B	(4) Treatment-H	(5) Survey Responders
Female	0.209	0.202	0.219	0.209	0.191
	(0.407)	(0.402)	(0.414)	(0.407)	(0.394)
Black	0.0527	0.0545	0.0560	0.0386	0.0273
	(0.223)	(0.227)	(0.230)	(0.193)	(0.163)
Hispanic	0.0812	0.0864	0.0807	0.0643	0.0747
	(0.273)	(0.281)	(0.273)	(0.246)	(0.263)
Asian	0.0854	0.0817	0.0846	0.0997	0.0893
	(0.279)	(0.274)	(0.279)	(0.300)	(0.285)
Verbal SAT/100	6.333	6.333	6.396	6.178	6.436
	(0.666)	(0.664)	(0.739)	(0.404)	(0.683)
Math SAT/100	6.589	6.578	6.564	6.691	6.753
	(0.648)	(0.645)	(0.680)	(0.565)	(0.627)
Recruited Athlete	0.218	0.222	0.232	0.174	0.142
	(0.413)	(0.416)	(0.422)	(0.379)	(0.349)
Attended Prep School	0.170	0.166	0.186	0.141	0.128
	(0.376)	(0.372)	(0.390)	(0.349)	(0.334)
Academic Ent. Score	0.0595	0.0663	-0.123	0.487	0.196
	(0.983)	(0.982)	(0.998)	(0.800)	(0.951)
Leadership Ent. Score	0.0267	0.0179	0.0294	0.0500	0.110
	(0.983)	(0.973)	(0.993)	(0.997)	(1.004)
Fitness Ent. Score	0.0350	0.0414	0.206	-0.408	-0.00339
	(0.979)	(0.967)	(0.971)	(0.902)	(0.988)
Sophomore GPA	2.835	2.821	2.812	2.941	2.991
	(0.547)	(0.551)	(0.562)	(0.482)	(0.534)
High Verbal	0.255	0.252	0.358	0.0129	0.311
	(0.436)	(0.434)	(0.480)	(0.113)	(0.464)
Other	0.463	0.468	0.251	0.968	0.499
	(0.499)	(0.499)	(0.434)	(0.177)	(0.500)
Predicted Low	0.305	0.303	0.423	0.0193	0.209
	(0.461)	(0.460)	(0.494)	(0.138)	(0.407)
Observations	2144	1065	768	311	549

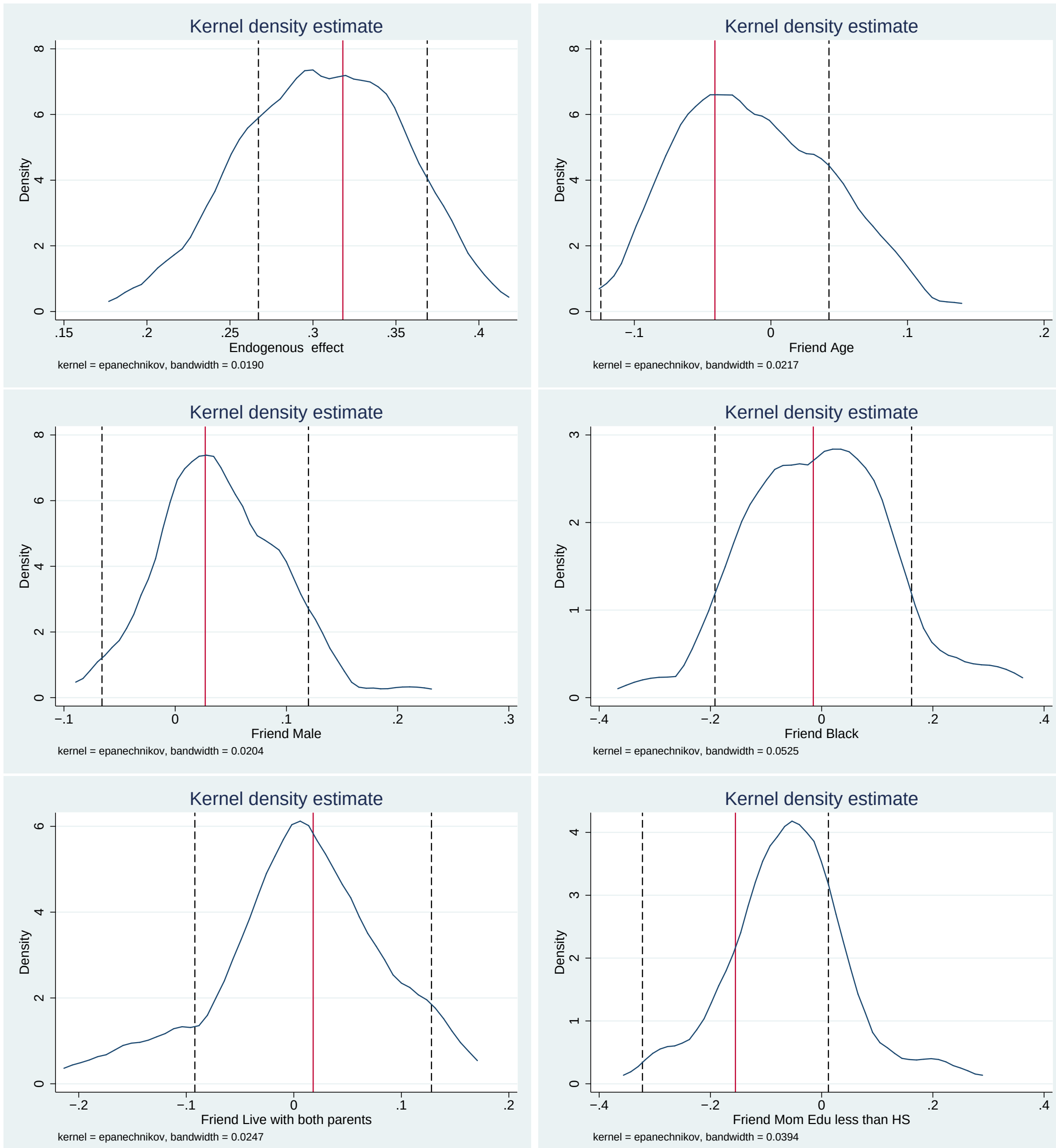
Means and standard deviations (in parentheses). Control refers to students who were assigned squadrons using an established stratified randomization. Students in the treatment group were assigned to squadrons in a way to increase achievement of predicted low-ability students (See Carrell et al. (2013)). This resulted in two categories of treatment squadrons: Bifurcated and Homogenous. Bifurcated squadrons were made up of largely high-ability and low-ability students while Homogenous squadrons were made up of largely middle-ability ones.

Table 2: Social Network Survey Summary

	Any	Study	Friend
Response Rate	0.256	0.234	0.248
Responders	549	501	531
Average Listed	2.727	1.756	1.905
Reciprocity Rate	0.464	0.557	0.594
Percent Listed	0.670	0.550	0.600
Observations	2144	2144	2144

Students were asked to name up to 5 friend and 5 study partners from their sophomore year. Response rate was 25%. The data above represent averages from the survey, except for Percent Listed. Percent Listed gives the number of USAFA students in the data (2144) who either responded or were named on at leat one survey.

Figure A1: Network Imputation Validation Using the Add Health Data Set



Red line - Lin estimate
Density plot - coefficients from 50 repetitions, starting from 25% sample
Black lines - 90% CIs.

Figures represent distributions of estimates of Lin model estimated in Appendix B. For each graph, the solid vertical line represents a point estimate obtained using the full Add Health data set. Vertical dashed lines represent the 90% confidence interval of that estimate. The distribution represents the set of point estimates obtained when performing network reconstruction, starting with 25% of Add Health friendship nominations and simulating the rest.

Table 3: Friendship Nominations Patterns

	(1) Connected	(2) Study	(3) Friends
Female-Female	0.710***	0.726***	0.625***
	(0.136)	(0.152)	(0.154)
Female-Male	-1.013***	-0.917***	-1.398***
	(0.0850)	(0.0913)	(0.113)
Athlete-Athlete	0.0387	0.234	-0.197
	(0.181)	(0.201)	(0.188)
Athlete-Not	-0.606***	-0.531***	-0.732***
	(0.0799)	(0.0876)	(0.0972)
Both Minority	-0.0653	-0.259	0.244
	(0.149)	(0.185)	(0.166)
Both White	0.0899	0.0118	0.177**
	(0.0658)	(0.0756)	(0.0789)
Low-Low	-0.0969	0.00995	-0.220
	(0.131)	(0.154)	(0.158)
Low-Med	-0.104	-0.152	0.000219
	(0.0908)	(0.110)	(0.104)
Low-High	-0.234***	-0.225**	-0.140
	(0.0863)	(0.103)	(0.105)
Med-Med	-0.00119	0.110	0.0409
	(0.0930)	(0.108)	(0.109)
Med-High	-0.252***	-0.232**	-0.245**
	(0.0803)	(0.0956)	(0.0962)
Constant	-1.608***	-2.054***	-2.046***
	(0.0776)	(0.0884)	(0.0955)
Observations	14369	14369	14369
Dep. Mean	.118	.079	.083

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Standard errors in parentheses. Standard errors are clustered by sophomore squadron assignment. Dependent variable is binary measure of whether or not student named other student in squadron as a friend or study partner. Columns 1 and 2 combine friend and study partner nomination data. Column 3 reports only study partner results. Column 4 reports only friendship nominations. Columns 1, 3 and 4 assume friendships are mutual, Column 2 assumes the nominator is a fried of the nominee, but the nominee does not reciprocate.

Table 4: Impact of Experimental Treatment on Sophomore Year Grades

	(1) First Stage
Instruments	
Treat X Low	-0.0608*
	(0.0329)
Treat X Mid	0.0705**
	(0.0258)
Treat X Top	0.0692*
	(0.0409)
Experimental Controls	
GPA Bot	-0.461***
	(0.0409)
GPA Mid	-0.264***
	(0.0245)
High Verbal	0.0451
	(0.0417)
F-Stat	7.463
Prob > F	0.0002
N	2134
R-squared	0.325
Indv. Controls	Y
Peer Controls	Y

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Standard errors in parentheses. Standard errors are clustered by freshman squadron assignment. Dependent variable is sophomore year GPA. Bi x Low (Bi x High) indicate a student who was predicted to be in the bottom (top) tercile academically and was in a bifurcated treatment squadron. Control students are the omitted treatment status.

Table 5: Preferred Estimates, Using Experimental Squadron Status as Instrument For Friend GPA

	(1) OLS	(2) IV	(3) Context	(4) Full
Contemporaneous Effect	0.138	0.101		0.107
	[0.03]**	[0.05]*		[0.05]**
	(0.03)	(0.04)		(0.04)
Contextual Effects				
Friend SAT V			-0.022	-0.023
			[0.07]	[0.08]
			(0.02)	(0.02)
Friend SAT M			-0.088	-0.123
			[0.06]	[0.07]*
			(0.02)	(0.02)
Friend ACA			-0.110	-0.124
			[0.04]**	[0.05]**
			(0.02)	(0.01)
Friend Prep			0.226	0.241
			[0.13]*	[0.14]*
			(0.04)	(0.04)
Friend Black			-0.610	-0.655
			[0.25]**	[0.27]**
			(0.08)	(0.08)
Friend Hisp			0.231	0.217
			[0.22]	[0.24]
			(0.08)	(0.07)
Friend Asian			-0.088	-0.123
			[0.71]	[1.63]
			(0.02)	(0.02)
Friend Female			-0.048	-0.121
			[0.20]	[0.22]
			(0.07)	(0.07)
N	2134.1	2135.3	2144	2135.3
Avg IV F-Stat		7.46		7.46
Repetitions	250	250	1	250
Contextual Controls	N	N	Y	Y
Indv Controls	Y	Y	Y	Y

Dependent variable is sophomore year GPA. Results from Columns 2 and 4 represent averages over 250 repetitions of simulated social network reconstruction. Average t-statistic in parentheses, standard deviation of point estimates in brackets. Standard errors are clustered by sophomore squadron.

Table 6: Preferred Estimates, Full Coefficient List

		(1) Full	
Contemporaneous Effect	0.107	[0.05]**	(0.04)
Contextual Effects			
Friend ACA	-0.124	[0.05]**	(0.01)
Friend SAT V	-0.023	[0.08]	(0.02)
Friend SAT M	-0.123	[0.07]*	(0.02)
Friend Ldrship	0.031	[0.05]	(0.01)
Friend CFT	0.015	[0.05]	(0.02)
Friend Ath	-0.002	[0.14]	(0.06)
Friend Prep	0.241	[0.14]*	(0.04)
Friend Black	-0.655	[0.27]**	(0.08)
Friend Hisp	0.217	[0.24]	(0.07)
Friend Asian	-0.123	[1.63]	(0.02)
Friend Female	-0.121	[0.22]	(0.07)
Individual Controls			
ACA	0.192	[0.01]***	(0.00)
SAT V	0.124	[0.01]***	(0.00)
SAT M	0.145	[0.01]***	(0.00)
Leadership	-0.001	[0.01]	(0.00)
CFT	0.056	[0.01]***	(0.00)
Athlete	0.055	[0.03]**	(0.01)
Prep School	-0.145	[0.03]***	(0.01)
Black	-0.010	[0.04]	(0.01)
Hispanic	-0.001	[0.04]	(0.01)
Asian	-0.000	[0.04]	(0.01)
Female	0.009	[0.03]	(0.01)
Treat-B as Fr	-0.058	[0.03]*	(0.01)
Treat-H as Fr	0.074	[0.03]**	(0.01)
N	2144		
Repetitions	250		

Dependent variable is sophomore year GPA. Results represent averages over 250 repetitions of simulated social network reconstruction. Average t-statistic in parentheses, standard deviation of point estimates in brackets. Standard errors are clustered by sophomore squadron.

Table 7: Distinguishing Between Different Types of Social Connections

	(1) Both	(2) Study	(3) Friend	(4) Directed
Contemporaneous Effect	0.108	0.133	0.043	0.087
	[0.05]**	[0.05]**	[0.05]	[0.05]*
	(0.04)	(0.04)	(0.04)	(0.04)
Contextual Effects				
Friend ACA	-0.115	-0.110	-0.106	-0.111
	[0.05]**	[0.05]**	[0.05]**	[0.05]**
	(0.01)	(0.01)	(0.01)	(0.01)
Friend Prep	0.245	0.278	0.271	0.256
	[0.13]*	[0.13]**	[0.13]**	[0.13]*
	(0.03)	(0.04)	(0.03)	(0.03)
Friend Black	-0.593	-0.555	-0.599	-0.598
	[0.25]**	[0.26]**	[0.26]**	[0.25]**
	(0.04)	(0.05)	(0.05)	(0.04)
Friend Hisp	0.203	0.181	0.172	0.207
	[0.20]	[0.20]	[0.20]	[0.20]
	(0.03)	(0.05)	(0.05)	(0.03)
Friend Asian	-0.098	-0.083	-0.062	-0.086
	[0.22]	[0.22]	[0.22]	[0.22]
	(0.02)	(0.02)	(0.02)	(0.02)
Friend Female	-0.070	-0.054	-0.024	-0.064
	[0.55]	[0.31]	[0.30]	[0.43]
	(0.04)	(0.06)	(0.07)	(0.04)
N	2135.3	2137.1	2137.6	2135.8
Avg IV F-Stat	7.46	7.46	7.46	7.46
Repetitions	250	50	50	50
Contextual Controls	Y	Y	Y	Y
Indv Controls	Y	Y	Y	Y

Dependent variable is sophomore year GPA. Results represent averages over multiple repetitions of simulated social network reconstruction. Average t-statistic in parentheses, standard deviation of point estimates in brackets. Column 1 treats study partners and friends equally. Column 2 uses only information on study partner connections. Column 3 only uses information on friendships. Column 4 uses both study partner and friendship information, but treats friendships as directed (i.e. non-reciprocal. Standard errors are clustered by sophomore squadron.)

Table 8: Alternate Instrument: Friends-of-Friends

	(1) Both	(2) Study	(3) Friend	(4) Directed
Contemporaneous Effect	0.299	0.304	0.230	0.190
	[0.38]	[0.32]	[0.40]	[0.35]
	(0.40)	(0.36)	(0.42)	(0.37)
Contextual Effects				
Friend ACA	-0.051	-0.029	-0.040	-0.031
	[0.07]	[0.06]	[0.08]	[0.07]
	(0.08)	(0.07)	(0.08)	(0.07)
Friend Prep	0.012	-0.005	-0.000	0.020
	[0.07]	[0.15]	[0.01]	[0.06]
	(0.07)	(0.07)	(0.08)	(0.05)
Friend Black	-0.002	-0.018	0.087	-0.000
	[0.06]	[0.08]	[0.09]	[0.02]
	(0.08)	(0.06)	(0.08)	(0.07)
Friend Hisp	0.001	-0.018	0.019	0.021
	[0.04]	[0.06]	[0.05]	[0.05]
	(0.05)	(0.04)	(0.05)	(0.05)
Friend Asian	-0.029	-0.012	-0.035	-0.011
	[-0.05]	[-0.10]	[-0.02]	[-0.03]
	(0.06)	(0.05)	(0.06)	(0.06)
Friend Female	-0.034	-0.033	-0.037	-0.031
	[0.05]	[0.05]	[0.05]	[0.04]
	(0.04)	(0.04)	(0.03)	(0.04)
N	1961.5	1967.8	1966.8	1934.1
Repetitions	250	50	50	50
Friend Contextual Controls	Y	Y	Y	Y
Indv Controls	Y	Y	Y	Y

Dependent variable is sophomore year GPA. Results represent averages over multiple repetitions of simulated social network reconstruction. Average t-statistic in parentheses, standard deviation of point estimates in brackets. Standard errors are clustered by sophomore squadron.

Table 9: Weighted Based on Predicted Social Connections

	(1)	(2)	(3)	(4)
	Both	Study	Friend	Directed
Contemporaneous Effect	0.120	0.137	0.065	0.098
	[2.21]**	[2.40]**	[1.15]	[1.87]*
	(0.04)	(0.05)	(0.04)	(0.04)
Contextual Effects				
Friend ACA	-0.131	-0.127	-0.128	-0.132
	[-2.61]**	[-2.23]**	[-2.19]**	[-2.57]**
	(0.01)	(0.02)	(0.03)	(0.02)
Friend Prep	0.269	0.304	0.313	0.288
	[1.78]*	[1.83]*	[1.79]*	[1.82]*
	(0.04)	(0.06)	(0.08)	(0.06)
Friend Black	-0.635	-0.658	-0.646	-0.634
	[-2.27]**	[-2.07]**	[-1.97]*	[-2.17]**
	(0.08)	(0.12)	(0.15)	(0.12)
Friend Hisp	0.213	0.251	0.224	0.221
	[0.89]	[0.96]	[0.81]	[0.88]
	(0.07)	(0.11)	(0.14)	(0.09)
Friend Female	-0.163	-0.192	-0.165	-0.164
	[-0.70]	[-0.75]	[-0.60]	[-0.67]
	(0.06)	(0.10)	(0.11)	(0.08)
N	2135.3	2137.1	2137.6	2135.8
Avg IV F-Stat	7.46	7.46	7.46	7.46
Repetitions	250	50	50	50
Peer Controls	Y	Y	Y	Y
Indv Controls	Y	Y	Y	Y

Dependent variable is sophomore year GPA. Results represent averages over multiple repetitions of simulated social network reconstruction. Average t-statistic in parentheses, standard deviation of point estimates in brackets. Standard errors are clustered by sophomore squadron.

Table A1: Selection into Survey Response

	(1)
	Survey Taker
Survey Taker	
Verbal SAT	0.0437
	(0.0949)
Math SAT	0.396***
	(0.0947)
Leadership Composite Entry	0.0534*
	(0.0309)
Fitness Entry Score	0.0120
	(0.0758)
Academic Composite Entry	0.0441*
	(0.0261)
Recruited Athlete	-0.433***
	(0.147)
Attended Prep School	-0.0731
	(0.171)
Black	-0.396
	(0.295)
Hispanic	-0.0186
	(0.196)
Asian	-0.00872
	(0.180)
Female	-0.0557
	(0.132)
Homogenous	0.224
	(0.150)
Bifurcated	0.0387
	(0.113)
Constant	-5.447***
	(1.011)
Observations	2144

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Standard errors in parentheses. Standard errors are clustered by sophomore squadron assignment. Dependent variable is binary measure of whether or not student filled out a friendship survey.

Table A2: Triad Formation Patterns

	(1) Pct of Pop	(2) Freq	(3) Vs. Random
All Female	0.00679	0.0467	20.18
Mix Gender, High-Low	0.221	0.00200	0.864
Mix Gender, No H-L	0.293	0.00283	1.222
All Male, H-L, All White	0.0928	0.00664	2.869
All Male, H-L, Mix Race	0.117	0.0114	4.910
All Male, No H-L, All White	0.134	0.0138	5.947
All Male, No H-L, Mix Race	0.137	0.0110	4.747
Observations	1410444	1410444	1410444

Shows the different triad group probability estimates. Column 1 shows the perctange of each group triad type in the whole population of triads (N=235074). Column 2 shows, by group, the frequency of total triads that are complete (i.e. where all three students are friends). This was measured using only triads where at least 2 of the respondents were completed the survey. Column 3 shows how the frequency from column 2 compares to a baseline where all links are independent of each other. Baseline was calculated using the cube of average frequency of dyads among survey takers.

Table B1: Network Reconstruction Simulation

	(1) Reconstructed	(2) Full Response	(3) Full Sample
Endogenous effect	0.305	0.318	0.274
	[0.039]***	[0.031]***	[0.005]***
	(0.046)		
Own Age	-0.202	-0.187	-0.122
	[0.030]***	[0.029]***	[0.004]***
	(0.012)		
Own Male	-0.104	-0.092	-0.175
	[0.035]***	[0.036]***	[0.006]***
	(0.014)		
Own Black	-0.102	-0.040	-0.161
	[0.065]*	[0.082]	[0.012]***
	(0.024)		
Own Asian	0.186	0.176	0.221
	[0.111]**	[0.110]*	[0.014]***
	(0.035)		
Own Live with both parents	0.086	0.078	0.124
	[0.041]**	[0.040]**	[0.007]***
	(0.013)		
Own Mom Edu less than HS	-0.180	-0.093	-0.078
	[0.062]***	[0.060]*	[0.010]***
	(0.027)		
Friend Age	-0.012	-0.041	-0.046
	[0.049]	[0.051]	[0.008]***
	(0.053)		
Friend Male	0.041	0.027	0.060
	[0.055]	[0.056]	[0.010]***
	(0.054)		
Friend Black	-0.003	-0.015	-0.037
	[0.107]	[0.107]	[0.017]**
	(0.128)		
Friend Asian	0.053	0.103	0.056
	[0.196]	[0.198]	[0.022]***
	(0.187)		
Friend Live with both parents	0.006	0.018	0.080
	[0.067]	[0.067]	[0.012]***
	(0.076)		
Friend Mom Edu less than HS	-0.057	-0.155	-0.021
	[0.103]	[0.101]*	[0.017]
	(0.107)		
N	1608.5	1621	67163
Repetitions	50	1	1

Each column estimates MLE model described in Lin (2010) using the Addhealth data set. Column 1 uses a 1,621 student subsample of the Addhealth data set. 25% of true Addhealth friendship nominations were used and the full network was reconstructed in 50 separate repetitions. Average standard errors are in square brackets and the standard deviation among point estimates is below in parentheses. Column 2 gives the results on the same subsample when 100% of true friendship nominations were used. Column 3 shows results for the full Addhealth population of 67,163 students and matches model 6 from Lin (2010). Stars represent significance at the 10%, 5% and 1% levels. Additional own and friend characteristics estimated, but not reported include: Years in school, Hispanic, other race, plays sports, mother education more than HS, mother education missing, mother working, mother on welfare and mother information missing.