Does who you know matter? Unraveling the influence of student networks on academic performance

Tarun Jain* Nishtha Langer†

October 27, 2013

Abstract

This paper examines the impact of students' network size and distance on academic performance. Larger and closer networks facilitate information exchange and knowledge appropriation, but may also reduce match-specific synergies that decrease productivity. Network data from a business school where students are randomly assigned to multiple overlapping sets of peers allows us to calculate centrality measures. Increasing closeness centrality within the network has a negative effect on student performance as measured by grade point average, suggesting that synergy reduction and information processing costs outweigh benefits from greater information access. In contrast, increasing ability among direct connections positively affects academic performance.

Keywords: Peer effects. Social networks. Management education.
JEL Codes: I23, L23.

*Email: tj9d@virginia.edu. Assistant Professor of Economics and Public Policy, Indian School of Business, Hyderabad AP India.

†Email: nishtha_langer@isb.edu. Assistant Professor of Information Systems, Indian School of Business, Hyderabad AP India. We thank Ashalata Devi, Kishore H., V. Srinath and Meenakshi Devi for assistance in compiling the administrative data and Lalita Reddi for outstanding research assistance. This paper benefited from detailed feedback from Amit Bubna, Maitreesh Ghatak and John Leahy. We thank seminar and conference participants at IIT Kanpur, ISB, University of Connecticut, Econometric Society Asian Meetings, ISIS 2013, SCECR 2013 and WISE 2012 for helpful comments on this paper. All errors are our own.

1 Introduction

A rich literature in economics establishes that the social environment is important for human capital accumulation and productivity. In particular, peers have significant influence on productivity in academic and workplace settings through mechanisms such as learning, specialization and competition.[1] The focus of most papers is on dyadic relationships in relatively small groups, which is helpful in understanding the precise impact of immediate peers. However, dyadic relationships miss key dynamics of the social environment. Specifically, the value that peers bring to relationships might be shaped by others they are connected to. Less is known about how these extended networks beyond the immediate peers affect individual productivity. Mapping the entire social network can account for richer relationship structures (Gulati 1999), and offers the ability to analyze the impact of an individual's network size and distance on productivity.

We analyze the impact of network variables such as the number of immediate connections (degree centrality) and the distance of an individual to other individuals (closeness centrality) on academic outcomes of students at a business school.[2] We consider the influence of network size and distance on human capital achievement, which is important both as a measure as well as a determinant of productivity. We examine (i) the effect of varying the number of network connections and the relative distance to other students within the network on academic performance, (ii) the effect of student ability combined with network measures on grades, (iii) the effect of the size, distance, and ability of immediate network connections on academic achievement, and (iv) heterogenous impact of network measures on different types of students. We find that the extended network has a significant effect on productivity, with lower distance decreasing productivity, especially as student ability increases, suggesting that synergy reduction and information processing

[1]The literature has examined peer effects in academic (Sacerdote 2001; Foster 2006; Stinebrickner and Stinebrickner 2006; Carrell, Fullerton, and West 2009; Jain and Kapoor 2012) as well as workplace settings (Mas and Moretti 2009; Guryan, Kroft, and Notowidigdo 2009; Nanda and Sørensen 2010). Lerner and Malmendier (2012) and Shue (2013) estimate the impact of business school peers on entrepreneurship, executive compensation and firm performance.

[2]We do not examine other network centrality measures such as page rank and eigenvector centrality because these measures are highly correlated with degree or closeness centrality. We do not use betweenness centrality because the network has relatively fewer structural "holes" so the data do not show much variation for this measure.

costs outweigh benefits from greater information access. In contrast, we find that increasing ability among direct connections positively affects academic performance.

The current theoretical literature is ambivalent on the value individuals derive from being part of a network and has argued for both the benefits (Jackson and Wolinsky 1996; Jackson and Rogers 2005) as well as costs (Jackson and Wolinsky 1996; Borgatti and Cross 2003) of peer networks. For instance, there are a number of reasons why the number of network connections and distance to other students might positively influence student productivity. Consider the connections model (Jackson and Wolinsky 1996) in which agents benefit from more links to other agents, with closer connections more beneficial than distant ones. In the context of peer networks in a business school setting, students may rely on other students to learn academic concepts. Larger networks may thus facilitate information accumulation from a wider set of peers, aiding students' understanding of different concepts and therefore enhancing productivity. Even when the benefits of connections drop off disproportionately with distance, for instance if the accuracy of information from indirect connections is lower (the "small worlds" formation of Jackson and Rogers 2005), larger networks may still be beneficial.

In addition to a student's own network connections and distance, the characteristics of immediate peers might affect productivity as well. Many workers learn about job openings through social connections (Calvo-Armengol and Jackson 2004, Calvo-Armengol and Jackson 2007). Since a major motivation for student to attend business school is to improve career trajectories and outcomes, better employment status of a student's connections might increase the likelihood that those connections will share information on job openings with the focal student. Recent empirical evidence supports positive network effects. For instance, Aral and Walker (2013) conduct an experiment on Facebook to promote positive social change and find that greater closeness increases influence over peers. In a study on microfinance conducted in rural Indonesia, Alatas et al. (2012) find that individuals with more connections are also better informed about the financial holdings of their peers.

At the same time, networks can also impose negative costs in the knowledge context. In Jackson and Wolinsky (1996)'s co-author model, match-specific synergies imply that every additional connection is harmful because interaction with a larger number of peers requires effort that reduces disproportionately the benefit from each link. Similarly, more central individuals might incur greater costs while searching and sorting to find the most relevant information, which suggests that closer ties do not necessarily improve productivity (Borgatti and Cross 2003). This study seeks to resolve the ambiguous impact of network connections and relative distance on productivity.

A number of challenges are associated with empirical research on network effects. First, researchers rarely have information about the complete network structure. Even with random sampling, sampled network nodes might be systematically different from unsampled nodes. Additionally, even if similar to sampled nodes, missing connections lead to an incorrect understanding of the underlying network structure and biased estimates of network effects (Chandrasekhar and Lewis 2011). Second, even when complete information about the network structure is available, network effects may be confounded with other endogenous effects (Manski 1993). Networks tend to form between people who believe the association will be beneficial, so self-selection may be difficult to identify separately from network effects. Connections also tend to form between individuals exhibiting homophily so students who link to each other or are close to each other in a network may share similar attributes, associate with each other due to mutually shared interests or may be influenced by shared environmental characteristics. Thus, the effects of homophily may be incorrectly attributed to network effects.[3] These factors make it challenging to cleanly identify the effect of network characteristics on productivity separately from the impact of self-selection into networks or correlated environmental effects.

This paper uses administrative data from a business school setting where students are exogenously assigned to interlinked networks of different sizes, and where the structure of mutually exclusive interlocking connections allows mapping the entire student cohort in a single network. This not only allows clean identification of network effects, but also addresses concerns with selective

[3] For instance, Ductor et al. (2013) who examine the role of coauthor networks on research productivity cannot exclude the role of homophily or strategic behavior coauthorship formation. Aral, Muchnik, and Sundararajan (2009), Aral and Walker (2011), and Shalizi and Thomas (2011) discuss the implications of homophily in detail.

sampling of network nodes. Our network measures examine the impact of network size and relative distance on well defined and objective measures of human capital formation and productivity. Thus, this paper is able to overcome significant empirical challenges to produce clean and unbiased estimates of peer network effects. While a number of studies based on undergraduate students find relatively small influence of peers (Sacerdote 2001; Foster 2006; Stinebrickner and Stinebrickner 2006), our study follows the work of Jain and Kapoor (2012), Lerner and Malmendier (2012) and Shue (2013) by examining productivity in a setting where networks are expected to be influential.

We find that student performance is statistically uncorrelated with degree or closeness centrality, in contrast to Aral and Walker (2013) and Alatas et al. (2012) who report that more connections are associated with greater productivity. However, our results show that closeness centrality has a negative effect with increasing ability (as measured by GMAT scores), especially among students with below median GMAT scores, lending support to the co-author model, where more connections for peers lower the value of those connections and reduce productivity.

Connections to highly connected individuals have no significant effect on performance. However, the effect of maximum ability among an individual's direct connections is approximately one-fourth the effect of own ability on productivity and statistically significant. Thus, it is not who you know, but what they know that is important.

We also examine impact heterogeneity, i.e., whether different types of students, specifically high versus low ability students and male versus female students, experience the influence of network measures on academic achievement differently. Our results show that connections are costlier for high ability students compared to low ability students, but no major differences of network measures between men compared to women.

This study makes important contributions to the existing literature on the effects of peer networks in an academic setting: (i) we are one of the first to examine the impact of both direct and extended connections in academic peer networks on objectively measured performance outcomes, (ii) we explore how student heterogeneity moderates the effect of the network attributes, (iii) our data and research setting allow us to specifically address the endogeneity and sampling concerns typical in research on network effects, and (iv) we extend the literature on strategic network formation to empirically test the predictions of the connections and the co-author models (Jackson and Wolinsky 1996).

These findings, which highlight the value as well as costs of network connections, are important not only for the education sector that we analyze, but also for other settings where individuals process information in a networked environment. Thus, we argue that lowering transaction and interaction costs is important while designing systems to disseminate information and knowledge within networks.

The rest of the paper is organized as follows. Section 2 introduces the institutional setting where the study is located and the data elements we use. We describe our dataset, discuss the assignment process that is the heart of our empirical model, and provide evidence to support our identification strategy. Section 3 analyzes these data in detail, including a discussion of the results, the impact of heterogeneity, and robustness checks. Section 4 concludes with a discussion of the policy implications and some suggestions for future work.

2 Data description and setting

Estimating the impact of peer networks on productivity requires micro-data from a setting where individuals are assigned exogenously to the network. The network should be complete and compact, with negligible impact of individuals outside the network on productivity. In order to estimate the impact of the network variables, the nodes should have exogenous variation in the number of connections. Finally, the dataset should contain well-defined measures of productivity such as grades and earnings, as well as a rich set of covariates that describe each student's ability, skills, professional background, and demographic characteristics. The next three sections describe the data that satisfy these requirements, and allow for estimation of the size and direction of peer effects.

2.1 Data source

The source of our data is the flagship post-graduate business program at the Indian School of Business (ISB). Established in 2001, ISB is a large, independent provider of post-graduate management education with a one year, full-time, residential diploma program. The school was founded in academic collaboration with the Wharton School at the University of Pennsylvania, Kellogg School of Management at Northwestern University, and London Business School, and incorporates many academic features and policies from its partner institutions.

An application to ISB consists of GMAT scores, essays, letters of recommendation, undergraduate and graduate transcripts, and an interview. Although drawing from a pool of applicants predominantly from India, Table 1 shows that student characteristics at ISB are comparable with those at a number of leading international business schools. Hence, ISB is arguably similar to a number of major international business schools on observable characteristics.[4]

Classes at ISB are held for 50 weeks without any significant break and are divided into eight terms of six weeks each. In the first four terms, students take a common "core" of 16 non-elective classes covering a range of management topics. In the next four terms, students choose various elective courses that allow them to concentrate in the areas of entrepreneurship, finance, information management, operations, marketing or strategy.

Instructors at ISB award course grades on a four-point scale. The highest grade is an A, corresponding to 4 grade points. Below this are A- (3.5 grade points), B (3 points), B- (2.5 points), C (2 points), D (1 point) and F (0 points). An F is a failing grade which requires the student to repeat the course. Instructors are required to maintain a class grade point average (GPA) between 3.25 and 3.30 across all sections that they teach. While student achievement is assessed on relative performance, the comparison set is all students in the sections that an instructor teaches (typically, 280 students in four sections) and not the students within the study group or even within the sec-

[4] A number of factors, such as location in a developing country, might differentiate ISB from other major management schools. However, without sector-wide micro-data from a large number of international schools, the impact of location, institutional, or cultural factors that might be correlated with the impact of peers is difficult to estimate.

tion.[5] Thus, a student's objective is to earn the maximum score possible, regardless of the relative performance of the other members in the study or residential groups.

The Academic Services Administration (ASA) at ISB maintains detailed records on the courses that each student enrols in, the grades achieved in these courses as well as assignment of students to study groups and residential facilities. We obtained a complete record of all enrolled students for four academic years from 2007-08 to 2010-11. An advantage of selecting this period was the absence of major changes in the curriculum or administrative policies during this time.

Student assignment, coursework and grade data are supplemented with data from admissions records that contain each student's academic (undergraduate and graduate institutions and associated majors and GMAT scores), professional (sector and firm of employment, employment duration, earnings, and functional role) and demographic backgrounds (year of birth, gender, marital status, and citizenship). Also included are data from the on-campus placement process; specifically, we use the earnings associated with the job accepted by students at the end of the PGP program.

Table 2 summarizes a number of relevant variables in the dataset. Enrolled students have an average of 4.9 years of full time work experience before joining. Consistent with the BusinessWeek data, the mean GMAT score is 709. Nineteen percent of students hold a masters degree before enrolling for ISB's program. The top two undergraduate alma maters are Delhi University (15 percent) and the Indian Institutes of Technology (14 percent from all campuses). In demographic characteristics, 73 percent of students are single at an average age of 28.7 years. Twenty six percent of the students are women, and 96 percent are Indian citizens. The average salary drawn before enrolling at ISB was INR 997,000 whereas the average salary reported on graduation was above INR 1,401,000, corresponding to 41 percent increase in compensation after one year of study.[6]

[5]Students do not know the correspondence between the class score and letter grades which is determined at the end of the term.

[6]At the time of writing, $USD1 = INR55$. The post-program earnings were coded as INR 1 for those students who started their own entrepreneurial ventures.

2.2 Network formation

This dataset offers a number of features that makes it attractive for analyzing network effects on the academic performance of business school students. First, the administrative source of the data allows us to map the entire set of peer network for each student, and avoid potentially biased estimates due to partial sampling from the network (Chandrasekhar and Lewis 2011). Second, since all administrative records are mandated to be complete and truthful, self-reporting bias, measurement error, and missing data do not threaten our analysis. Finally, in the one year PGP program, attrition is negligible and student cohorts do not overlap.[7] Therefore, non-random attrition from the sample as well as serial correlation due to overlapping peers across years is not a significant concern.

A key feature of the data that allows for network analysis is that students are simultaneously and randomly assigned to multiple sets of mutually exclusive peers.[8] Before the start of classes, ASA assigns each student to a study group that is expected to work together to understand the coursework, as well as to complete specific group-based assignments.[9] The study group assignment is fixed for the duration of the four core terms. In the elective terms, students choose their own courses, and these might be different from those of chosen by their core term study group peers.

In assigning students to study groups, ASA relies only on observable characteristics of students, following two simple sequential rules.[10] First, groups are assigned either two women, or none at

[7] In the entire sample period, only three students joined the program but left before completion. Seven students interrupted the program and rejoined in the next cohort. Both student types are dropped from the sample.

[8] The data also suffer from a few shortcomings. First, we cannot estimate the impact of endogenous, informal networks that might be correlated with GPA. Second, since students who conduct their own job search do not report earnings to ISB, the placement data are incomplete. If, for example, students who conduct their own job search are more likely to rely on professional and social networks, then estimates of influence of peers on earnings at graduation might suffer from systematic biases. Finally, this dataset does not contain information on students' family characteristics such as caste, religion or parental education that are potentially important in determining educational achievement. However, these factors are unlikely to be correlated with the group assignment.

[9] The number of sections increased from six in the 2007-08 and 2008-09 class years to eight in 2009-10 as the school increased enrolment from 416 students in 2007-08 to 436 students in 2008-09 and 565 students in 2009-10 and 2010-11.

[10] One of the authors observed this process and verified that the staff member had only demographic information for each student during the assignment process.

all after which the groups are balanced in the number of engineers. Each group formed with these restrictions consists of either four or five students. With these assignments, the data contain 90 study groups in the 2007-08 and 2008-09 class years, and 120 groups in the 2009-10 and 2010-11 class years. In particular, note that ASA neither considers any measure potentially correlated with ability, such as GMAT scores, elite undergraduate college or Master's degree, while assigning students to groups, nor any characteristic that is unobservable to the researchers such as motivation or potential for interaction with peers. Hence, due to the administrative process, the assignment of individuals to groups is statistically random on unobservable characteristics.

In addition to the study group network, students are also linked to a peer network in the residential dormitories. Unlike many international business schools, all students at ISB are required to stay on campus in housing provided by the school over the program duration. Students stay in either single apartments or quads with four bedrooms that share a kitchen and living space. Single apartments are assigned to students with cohabiting family members or those with special needs, such as physical disability.[11] Two observable assignment rules are followed while assigning students to quads. First, each quad is single sex. Second, roommates cannot overlap with study group peers. Once assigned, students stay in the same quad throughout the eight terms. Although there are more apartments than quads, most students live in quads. In the sample, 1484 out of 1987 students live in shared residences.

The mutually exclusive and interlinked study group and quad assignments are at the heart of how the peer to peer network of all students in each class is formed. Each student is connected to other students in a student group, and most of those students are connected to other students in quads, and so on till all students within a cohort are connected within a single network. For example, Figure 1 shows that student $S0$ is connected to $S1$, $S2$ and $S3$ in the quad, and $S4$, $S5$, $S6$ and $S7$ in the study group. Students $S4$ and $S6$ are further connected to other students in their quads ($S8$, $S9$ and $S10$ for $S4$, and $S11$, $S12$ and $S13$ for $S6$), whereas $S5$ and $S7$ might live

[11] ISB does not solicit data on roommate preferences, so we avoid potential homophily influencing network connections.

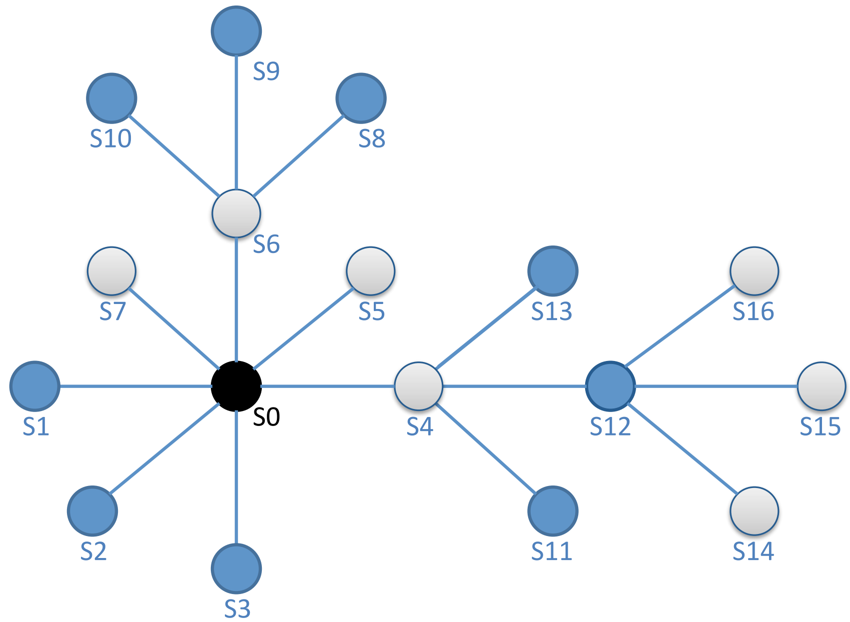

Figure 1: **Network example**

alone in single apartments.[12] Students might use these networks not only for completing study group submissions but also to reinforce their academic understanding and to discuss the learning objectives for a course (Jain and Kapoor 2012).

With these connections, we can construct four networks for each academic year, each of which includes all the students in the cohort, as shown in Figures 2. There is variation by student in the number of connections from two sources. First, as mentioned earlier, students might live in quads or single apartments, which will cause variation in the number of residential connections. Second, study groups comprise of either four or five students, which will cause variation in the number of study groups connections.

Using these variations, we calculate degree centrality, which for node i, is the number of direct connections and N is the total number of students in the cohort. Greater degree centrality for a student suggests that the student is able to aggregate more information from her peers, although she must consider the benefits and costs of information acquisition.

[12] The figure shows $S12$'s study group connections; although $S11$ and $S13$ are also in study groups, these connections are not shown.

$$degree_i = \sum_{j=1}^{N} A_{ij} \tag{1}$$

where $A_{ij} = 1$ if nodes i and j are directly connected and 0 otherwise. In Figure 1, the degree centrality of node $S0$ is 7, the number of direct connections.

We also calculate closeness centrality for each student. Closeness centrality is the average the shortest distance between a node and all other nodes. High closeness centrality implies that the other students in the network are relatively easy to reach, therefore, the student is likely more accessible to other nodes and hence ought to have greater influence in the broader network.

$$closeness_i = \frac{1}{\sum_{j=1}^{N} distance_{ij}} \tag{2}$$

where $distance_{ij}$ is the shortest path between nodes i and j. If the nodes shown in Figure 1 depicted a complete network, then the closeness centrality for node $S0$ would be 0.571, which is the inverse of the sum of the distances between $S0$ and each other node (28) divided by the number of other nodes (16). Table 2 reports that the mean degree centrality is 5.86, with a standard deviation of 1.34, and the mean closeness centrality is 0.00041 with a standard deviation of 0.00007.[13]

2.3 Randomization check

As we discussed earlier, a unique feature of this dataset that makes it appropriate for analysis of peer effects is that students are simultaneously and randomly assigned into two separate and mutually exclusive sets of peers – the study group and the residential group. Given the importance of random assignment in obtaining unbiased estimates, we check the effectiveness of the administrative process described above in the data. Even if the administrator did not account for ability in network creation, it is possible that the outcome of the process created groups with correlated ability. If that were the case, we might misattribute the impact of correlated ability to network effects.

[13]To ease interpretation and comparison of the regression coefficients, the empirical analysis in Section 3 uses z-scores associated with each of these metrics.

To check for random and uncorrelated assignment, we regress the mean GMAT scores of the direct connections on individual GMAT scores, including year dummies as control variables. To verify that the administrative process is also random with respect to an alternative measure of ability, we include a second set of regressions where mean pre-earnings of direct connections are regressed on individual pre-earnings. Since gender is the primary criteria for assignment of students to study groups and quads, we report results separately for women and men.

Table 3 shows that the correlation between a student's GMAT score and the mean GMAT scores for network connections is low. For both men and women, the coefficient associated with the study group's mean network GMAT scores (men: 0.023, women: 0.084) is smaller than the standard error (men: 0.054, women: 0.099). The pre-earnings test also reveals similarly that earnings before matriculation are uncorrelated within groups. The table shows that the key network measures used in the subsequent analysis, i.e., degree and closeness centrality, are uncorrelated with either GMAT or pre-earnings. These results support our claim that the administrative randomization process led to the formation of a network where ability is uncorrelated across nodes.

Our second data check ascertains that the assignment of students to single apartments is uncorrelated with ability and only dependent on demographic characteristics. According to housing department policies, single apartments are assigned when a student is accompanied on-campus by family members, most frequently spouse and children. Therefore, we test a model to empirically investigate the influence of various factors on whether a student is allocated housing in a single apartment versus a quad.

Table 4 reports both OLS coefficients and probit marginal effects associated with different factors that might potentially influence allocation. Both specifications consistently find that marital status is the most important factor that determines residence in a single apartment. OLS estimates in the first column show that singles are 56% less likely to allotted a single apartment. Men are more likely to have accompanying wives rather than vice versa due to cultural factors and are more likely to be in single apartments, as are international students. Other factors representing student ability such as GMAT scores, years of experience, pre-program earnings or a previous master's degree are uncorrelated with housing allocation.

The above analyses support our assertion that the students are allocated to the network randomly and not because of their ability which allows us to recover unbiased estimates of network effects. The next section describes the econometric analysis performed on this dataset to estimate the impact of peer networks on academic performance.

3 Empirical analysis

We consider a student's GPA from the core terms as the variable representing individual productivity. This measure has a number of advantages compared to alternative measures such as overall GPA or earnings after the program. First, all students simultaneously take the a set of non-elective "core" courses and cannot select the section, the instructor, or their peer study group. Second, students self-select into study groups in the elective courses, and hence selection effects cannot be distinguished from network effects. Third, the core terms grades are an objective measure of assessing student performance. Appendix A shows that even if GPA does not directly enter the utility function, students are motivated to maximize the core GPA because core term grades are strongly correlated with earnings after completing the program.

The following subsection presents the main empirical analysis, examining the impact of network measures on student GPA. The final two subsections present the results of further analysis that elucidate the heterogeneous impact of network interactions.

3.1 Impact of network on academic performance

We first present the model that estimates the impact of network measures on productivity. Given random allocation rules described earlier, identification of peer effects is not a significant concern. Although the dependent variable is theoretically truncated at 4.0 (the maximum GPA) and 0.0 (the minimum GPA), there are no observations at these points and OLS estimates will be consistent in reporting the impact of network characteristics on student outcomes. Therefore, we specify the following model where i represents a student, j is the set of direct connections to student i, and t represents a cohort.

$$\begin{aligned} y_{ijt} &= \beta_0 + \beta_1 \mathbf{X}_{ijt} + \beta_2 \mathbf{network}_{ijt} + \beta_3 GMAT_{ijt} * \mathbf{network}_{ijt} + \beta_4 mean(\mathbf{network}_{-ijt}) \\ &+ \beta_5 max(GMAT_{-ijt}) + year_t + \epsilon_{ijt} \end{aligned} \tag{3}$$

where $\mathbf{network}_{ijt} = [degree_{ijt}\ closeness_{ijt}]'$, $\beta_2 = [\beta_2^D\ \beta_2^C]$, $\beta_3 = [\beta_3^D\ \beta_3^C]$ and $\beta_4 = [\beta_4^D\ \beta_4^C]$. In this model, the main outcome of interest is the GPA for the core terms. Since individual characteristics directly impact productivity, the model includes a vector of variables, $\mathbf{X}_{ijt}$, representing own GMAT scores, years of full time experience, pre-program earnings and indicator variables for students who hold masters degree and who attended IIT and Delhi University for undergraduate studies. These variables help capture student maturity, experience with solving business problems, and success in the corporate world, perhaps representing individual drive and motivation. The vector $\mathbf{X}_{ijt}$ is augmented with demographic variables of the student such as gender, marital status and citizenship.[14]

The specification captures the influence of the peer network on individual performance through a number of variables. The first is degree centrality, which is the number of other students that an individual is directly connected ($degree_{ijt}$). The second is closeness centrality, which is the inverse of the sum of distances from an individual to all other students ($closeness_{ijt}$). We interact individual GMAT scores with network measures to represent variation in the impact of these network measures by student ability ($GMAT_{ijt} * \mathbf{network}_{ijt}$). If β_2^D is positive, then more direct connections for students benefit academic productivity consistent with the connections model. Conversely, $\beta_2^D < 0$ suggests that more direct connections are harmful to academic achievement, potentially due to greater opportunity cost of interaction. The coefficient β_2^C represents the effect of lower distance between a student and all other students in the network. However, since the specification

[14] Although the dataset contains information on student age, we exclude this variable from the specification since it is highly correlated with years of experience.

already controls for the number of direct connections, lower distance will raise closeness only if the number of connections for a student's direct connections increase. Therefore, $\beta_2^C < 0$ is consistent with the predictions of the co-author model, where more network connections for directly linked individuals harm productivity.

The coefficients β_3^D and β_3^C represent the marginal effect of degree and closeness centrality on students with increasing ability. β_3^D represents the marginal impact of greater degree and β_3^C is the effect of closer networks, both as students' GMAT scores increase. $\beta_3^D > 0$ suggests that more direct connections disproportionately benefit higher ability students, perhaps because they are better at processing information obtained from wider networks. $\beta_3^C > 0$ is consistent with closer connections disproportionately benefiting higher ability students again due to better information processing skills. Conversely, $\beta_3^C < 0$ indicates that closer connections disproportionately harm achievement among higher ability students. Since we have already controlled for GMAT interacted with degree centrality, β_3^C captures the effect of study group and roommates' connections rather than own connections on academic achievement.

The specification also includes variables representing characteristics of an individual's first degree connections, specifically the average degree centrality and average closeness among i's direct connections, and the maximum GMAT score among these direct connections. We interpret positive values for β_4^D and β_4^C to suggest that students benefit academically from better network connections and hence more information in the proximate network. Finally, β_5 is the impact of increase in ability of the individual with the highest GMAT score linked to student i. Thus, $\beta_5 > 0$ suggests that high ability peers have a positive impact on student achievement, consistent with the predictions of the models presented in Calvo-Armengol and Jackson (2004) and Calvo-Armengol and Jackson (2007).

The specification includes year fixed effects ($year_t$) to control for time-specific characteristics such as academic policies or macroeconomic conditions that are common within cohorts, as well as an i.i.d. error term to control for unobservable characteristics that might influence business school grades.

Column I in Table 5 reports the results from a model with only individual characteristics with subsequent columns adding network variables. Not surprisingly, an individual's own GMAT score is a statistically significant predictor of GPA, with one standard deviation increase in GMAT associated with 0.1 increase in core terms GPA. Pre-program salary, which may represent an individual's professional drive and motivation as well as job and industry-specific ability that is different from an academic ability captured by the GMAT score, is positively associated with GPA (+0.025 increase for one standard deviation increase in earnings, $p < 0.01$). We also find that women, single students older students, and international students report significantly lower grades.

Consistently across columns, we find that the effect of degree centrality on GPA is small and insignificant. Although closeness centrality is negatively associated with GPA, the coefficient is not robust since it is significant at the 5% level in Columns II and III, but loses statistical significance in Columns IV and V. These results indicate that the number of connections or relative distance to other students by themselves are not influential in predicting academic performance at business school.

The main finding from Columns III, IV and V is that increasing ability with lower mean distance is associated with lower productivity. Specifically, increasing $GMAT_{ijt} * degree_{ijt}$ by one standard deviation raises core terms GPA by 0.005 points, suggesting that the value of more direct connections increase as students are more capable of processing the information from these peers. However, the coefficient is imprecisely estimated in all three specifications. Conversely, increasing closeness centrality interacted with GMAT scores by one standard deviation leads to 0.033 point decline in core terms GPA ($p < 0.01$). As mentioned earlier, the specification already controls for the number of own direct network connections, both degree centrality and degree interacted with GMAT and therefore β_3^C represents the effect of connections for the extended network. The findings are consistent with a nuanced version of the co-author model, where more connections for direct connections are costly and lower individual productivity, particularly as a student's ability to use the information gleaned from networks increases.

At the same time, Table 5 reports relatively small and statistically insignificant coefficients associated with mean degree centrality and mean closeness centrality, suggesting that access to well-networked individuals do not affect productivity. In contrast, the impact of the maximum GMAT among direct connections is positive and statistically significant (+0.025 increase with one standard deviation in GMAT scores, $p < 0.01$). This coefficient is one-fourth the magnitude of own GMAT scores and comparable to the effect size of own pre-program earnings, suggesting that the value of peers increases in their ability.

Our main conclusion from the results presented in Table 5 is that while on average networks have an insignificant impact, close networks have a negative impact with increasing individual ability, offering support for the co-author model. Simultaneously, the effect of the maximum ability of the direct connection is large and significant, pointing to the salience of peer ability in determining productivity in an academic setting.

3.2 Heterogeneity in impact of network measures

This section examine heterogeneity in the results presented in Section 3.1. The strength of network effects might vary by student characteristics such as ability and gender. The impact of networks might also be sensitive to formal incentives for collective work. Finally, network effects might amplify over time as students learn to work with their connections, or diminish as students are able to study by themselves instead of relying on peers.

3.2.1 Heterogeneity in student characteristics

For many reasons, the strength of network effects uncovered in the previous section might be different for different types of students. For instance, students with low ability or professional experience might depend more on network connections for information exchange and learning, but may also be less efficient in allocating their time. Similarly, the results might vary by gender either if the structure or nature of social relationships with network connections is systematically different for men versus women.

To test this, Table 6 reports the results of analyzing equation (3) separately for high (with above median GMAT score) and low ability students. The effect of a student's GMAT score interacted with degree centrality is very different for high and low ability students. The coefficient associated with high ability students is negative (-0.052, $p < 0.05$) suggesting that such students experience greater opportunity costs of interaction with a wider set of immediate peers as GMAT scores increase. Conversely, the coefficient for low GMAT students is positive (+0.022, $p < 0.05$) which is consistent with greater academic benefits and lower opportunity costs of peer interaction as ability increases among such students.

The effect of closeness interacted with GMAT is also considerably different for high versus low ability students, with no impact on the former but a significant negative impact on the latter. One standard deviation increase in $GMAT_i * closeness_i$ for low ability students is associated with a 0.067 point decrease in GPA ($p < 0.01$). This finding suggest that the productivity loss when peers have larger networks is relatively little among the absolutely lowest ability students, but is increasing when these students gain information processing ability. Conversely, the null finding among high ability students suggests that the co-author effect has little productivity impact. This makes intuitive sense since such students are unlikely to need highly networked peers to improve their own productivity. In that sense, our findings echo Ductor, Fafchamps, Goyal, and van der Leij (2013), who also find that network characteristics of peers are no longer salient 14 years after the start of a research career.

Table 6 also reports that closeness centrality interacted with GMAT is associated with poorer performance for both men and women. However, the effect is stronger for women (-0.049, $p < 0.01$) than for men (-0.026, $p < 0.05$), with both coefficients statistically different from each other. This finding echoes Ibarra (1997) who found that women with greater advancement potential relied on close ties and relationships in decision making more than both high potential men as well as lower potential women. A possible explanation is that high ability women work more collaboratively, partly as a response to constraints imposed on them in gender-biased workplaces. Thus, more connections for collaborators of high ability women might lead to greater productivity loss compared to either high ability men or lower ability women.

These results encourage us to pursue a hierarchical Bayesian specification that allows us to exploit individual level heterogeneity (Rossi, McCulloch, and Allenby 1996; Rossi, Allenby, and McCulloch 2005), which if ignored can lead to biased coefficient estimates (Gonul and Srinivasan 1993; Gonul and Srinivasan 1996; Heckman 1981; Jain, Vilcassim, and Chintagunta 1994). For instance, the preceding analysis uncovers that ability and gender may affect how peer networks interact with student productivity. While we classified students on certain observable characteristics such as ability and gender to understand how these affect our model, we do not know *a priori* how heterogeneous observable and unobservable student attributes affect network measures and consequently, their productivity measures. To this end, we first re-specify our model as follows with core GPA as the outcome of interest.

$$\begin{aligned} y_{ijt} &= \beta_{0i} + \beta_{1i}\mathbf{X}_{ijt} + \beta_{2i}\mathbf{network}_{ijt} + \beta_3 GMAT_{ijt} * \mathbf{network}_{ijt} + \beta_{4i} mean(\mathbf{network}_{-ijt}) \\ &+ \beta_{5i} max(GMAT_{-ijt}) + year_t + \epsilon_{ijt} \end{aligned} \tag{4}$$

With this specification, we allow all coefficients to be student specific. This allows our model to cater to heterogeneous student attributes that affect ability, network, and control variables. The term ϵ_{ijt} represents unobservable factors that influence productivity and is assumed to be $\epsilon_{ijt} \sim N(0, \sigma_\epsilon^2)$.

We use gender and ability as the observable characteristics that contribute to heterogeneity that affect the magnitude of ability and network variables on student productivity and specify the following regression model.

$$\beta_{ijt} = \delta_0 + \delta_1 HighGMAT_{ijt} + \delta_2 Gender_{ijt} + \nu_{ijt} \tag{5}$$

where $\nu_{ijt} \sim i.i.d.N(0, V_\beta)$. The individual specific parameters are now functions of observable student attributes such as ability (above or below median GMAT score) and gender.

$$\beta_i = [\beta_{0i}\ \beta_{1i}\ \beta_{2i}\ \beta_{3i}\ \beta_{4i}\ \beta_{5i}\ \beta_{6i}] \tag{6}$$

In equation (5), the coefficients δ_1 and δ_2 denote how a student's ability and gender, respectively, affect the coefficients of the covariates specified in equation (4). For example, the effect of experience on δ_1 indicates the varying effect of degree centrality on GPA between students with above and below median GMAT scores. If $\delta_{1i} < 0$, then greater degree centrality is associated with lower GPA, this reduction in GPA is more for students with high GMAT scores. In other words, high ability students are shown to be more distracted by degree centrality than low ability students.

The random variable ν_i is an unobservable component of student heterogeneity, assumed to be distributed $N(0, \sigma_\nu)$. The associated variance-covariance matrix V_β determines the spread of the unobserved component. Using equation (5), we allow student attributes, both observable and unobservable, to affect our model parameters. Our intent is to demonstrate how student heterogeneity affects network variables that in turn affect productivity.

We thus take into account student heterogeneity and jointly estimate the model specified in equations (4) and (5) in a hierarchical Bayesian model. Our estimation process necessitates the computation of the exact information about the posterior distribution of the model parameters (Rossi, McCulloch, and Allenby 1996; Rossi, Allenby, and McCulloch 2005).

We follow standard estimation techniques for Bayesian models. We first set diffuse priors for the model parameters, then use Markov Chain Monte Carlo (MCMC) methods, specifically the Gibbs sampler, and data augmentation coded in R for our estimation. Such an approach is appropriate for the hierarchical structure inherent in our inference model. We build a Markov chain with a stationary distribution as the posterior. The estimation involves a series of draws till convergence of the posterior distribution is achieved. We run the MCMC simulation for 50,000 draws and discard the first 20,000 as burn in. We also use a thinning parameter of 20, i.e., we retain every twentieth of the remaining draws for the posterior distribution which helps in reducing storage space and alleviates the computational burden of analyzing stored draws (Rossi, Allenby, and McCulloch 2005).

We computed the regression coefficients for the sample, by averaging the posterior means of individual specific β_i's for each draw, these are similar to results in the full model (Table 7). In particular, we find that the coefficient associated with $closeness_{ijt}$ is negative and significant at 1% level, indicating that higher closeness centrality imposes opportunity costs that negatively affect academic performance. Likewise, consistent with our earlier results, the coefficient for closeness interacted with GMAT is also negative and significant (-0.019, $p < 0.01$).

We are, however, more interested in understanding how heterogeneity affects the model parameters. We present the estimations results of the posteriors distribution of the hierarchical regression coefficients from equation (5) in Table 8. First, we find that students manifest heterogeneity since the effects of ability and network variables on performance vary significantly across students. Specifically, the posterior mean for $Max(GMAT_{-i})$ in Table 7 is +0.023 ($p < 0.01$). The coefficients on $HighGMAT$ and $Gender$ for $Max(GMAT_{-i})$ indicate that while on average the maximum ability of the direct connections is positively correlated with own GPA, it is more positive when students are below the median GMAT or are female. It may be likely that these students are better able to interact or to derive greater value from their able and well-connected peers. These students perhaps have better time management skills or are able to absorb the opportunity costs posed by increased network interactions better than their peers.

We also find that while intrinsically degree is positively associated with GPA (+0.083, $p < 0.01$), this effect is smaller for women (-0.051, $p < 0.01$). Further, the interaction between GMAT and degree centrality is positive for women (+0.02, $p < 0.01$). In contrast, intrinsically, the direct effect of closeness centrality is negative (-0.472, $p < 0.01$), but again this effect is smaller for women. In addition, the interaction effect between GMAT and closeness centrality for women is negative. The literature suggests that high ability women rely disproportionately more on network connections (Ibarra 1997). Thus, decrease in attention from these connections might lead to greater productivity losses compared to either high ability men or lower ability women.

Overall, our results indicate that the effect of network variables and ability is different among different students. From a policy perspective, these results highlight the role administration may play in increasing the effect of peer to peer learning in large networks.

3.2.2 Heterogeneity in incentives

Core terms GPA, the outcome variable used in the previous section, consists of a component that varies with individual performance as well as a component that varies with the study group's joint performance. The structure of the data does not allow us to separate these components, so isolating the impact of the peers on the individual component of the GPA is difficult. Nonetheless, since the analysis presented in Section 3.1 does not show that study group peers have a statistically significant influence on individual GPA, we can perhaps conclude that mechanical correlation between individual and study group performance is not likely to drive the results.

In this section, we examine the impact of the size of group incentives in different courses on the strength of peer effects.[15] We separately estimate equation (3) on three outcome variables – GPA calculated for those courses where group work constitutes either high (35% to 50%), medium (20% to 30%) or low (0% to 15%) component of the final grade.

Table 9 shows that our main findings, that closeness is negatively associated with performance as ability increases, and that productivity increases with connections to high GMAT students, is maintained with all three outcome variables, suggesting that heterogeneity in formal incentives for group work are not major drivers for network interactions.

3.2.3 Heterogeneity over terms

Network effects may amplify over time as students learn how to seek information from peers. Conversely, they may dampen if students learn to work by themselves instead of relying from peers. For example, Ductor, Fafchamps, Goyal, and van der Leij (2013) find that the predictive power of coauthor networks lasts about fourteen years after the onset of the publishing career. This section investigates the impact of student networks on student GPA over each of the four core

[15]Note that the comparison between courses with a high versus low component of group work is not clean since the courses with more group work as part of the grade (typically in Strategy, Marketing, Entrepreneurship and Operations Management) are different from those with low group work (largely in Statistics, Economics and Finance).

terms.

Table 10 reports the impact of network measures from the full specification (equation 3) with each core term's GPA as the outcome variable. Most coefficients remain relatively stable over time, although we notice that the coefficient associated with the maximum peer GMAT is declining over time. While students continue to benefit from being connected to high ability individuals, they may become slightly more self-reliant over terms, with one student writing in a feedback survey completed at the end of the program that "the initial excitement to network was forced and short-lived." While the trend is consistent with the findings reported in Ductor, Fafchamps, Goyal, and van der Leij (2013) who also note the declining effect of network connections, network effects in Table 10 do not completely denude by the fourth term, perhaps because the time period considered is not long enough.

4 Conclusion

This paper investigates the impact of peer networks on productivity in an academic setting. A rich literature in economics and management finds that formal investments in human capital are important for improving productivity (Bapna, Langer, Mehra, Gopal, and Gupta 2013; Maitra and Mani 2012). The extant literature, especially the line of research with experimental or quasi-experimental assignment of peers pioneered by Sacerdote (2001), has identified how dyadic peer relationships affect performance outcomes. However, peer networks also have an important role in enabling information aggregation, learning and therefore, productivity (Gulati 1999). In particular, information exchange may be facilitated not only by direct ties an individual has, but also on where the individual is situated within the network. At the same time, an examination of the effect of networks on performance may be confounded by various factors. For instance, empirical analysis might be hindered if network formation is strategic (Jackson and Wolinsky 1996), networks manifest homophily or data is crippled by sampling errors (Manski 1993; Chandrasekhar and Lewis 2011).

Our paper contributes to the literature that has examined peer effects by focusing on the effect of network attributes such as degree and closeness centrality on academic outcomes. Using data from a business school setting that helps overcome significant empirical challenges, this paper uncovers the aspects of network structure that influence academic performance. By quantifying the effects of direct connections and network reach, our paper offers a richer perspective on how peers affect performance. Furthermore, by examining how these peer effects vary across students, our study affords a nuanced understanding of the co-authors model (Jackson and Wolinsky 1996). Our results provide empirical support for the co-author model, and indicate that connections for peers deter performance, presumably as those peers are distracted with increasing demands for their attention. These effects manifest more strongly for students with relatively low ability. In addition, the paper reports that academic performance improves with the maximum GMAT among the direct connections, or in essence, it is not who you know, but what they know.

Our research thus has both theoretical and practical significance, underlining not only the importance of connections to knowledgeable nodes, but also highlighting the dichotomy of being too connected. Our study implies that lowering the transaction and interaction costs is important while designing systems to disseminate information and knowledge within networks. In addition, our study highlights the perils of the co-author model, especially for low capability individuals. Finally, the measure of productivity used in the analysis is students' GPA in a business school. After controlling for various demographic and professional variables, GPA is a robust predictor of short-term earnings and we expect it to affect long-term earnings as well. Therefore, we expect that these findings have implications not only for performance in educational institutes but also for human capital formation and workplace productivity.

Our findings should be read with a few caveats. We examine academic outcomes measured while the network was in place, which leaves open the possibility that the results for the students in our sample might be very different when considering persistent academic or professional outcomes after the network assignments have ended.[16] Further, the networks we examine are exogenous, and

[16] A related issue is that we do not examine job search, promotion, or executive compensation outcomes that may be important in a business school setting.

hence we do not examine self-formed links between students who manifest homophily or believe that connections with other students might be beneficial. Insofar that endogenous relationships are formed and maintained only if beneficial, our results may represent a lower bound on the value of network connections. Readers are also cautioned while directly extending the coefficients estimated from the empirical exercise to other network situations, such as microfinance, job search or scientific research, where the impact of networks can be influenced by the environment and incentives for interaction as well as agents' ability to process information gained from the network.

Nonetheless, we hope that future research will link network effects and productivity in comprehensive models that can evaluate the outcome of specific policies (such as alternate network assignments or incentives for collaboration) on student outcomes.

References

Alatas, V., A. Banerjee, A. Chandrasekhar, R. Hanna, and B. Olken (2012). Network structure and the aggregation of information: Theory and evidence from Indonesia. NBER Working Paper No. 18351.

Aral, S., L. Muchnik, and A. Sundararajan (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences 106*(51), 21544–21549.

Aral, S. and D. Walker (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science 57*(9), 1623–1639.

Aral, S. and D. Walker (2013). Tie strength, embeddedness and social influence: Evidence from a large scale networked experiment. SSRN Working Paper No. 2197972.

Bapna, R., N. Langer, A. Mehra, R. Gopal, and A. Gupta (2013). Human capital investments and employee performance: An analysis of IT services industry. *Management Science 59*(3), 641–658.

Borgatti, S. and R. Cross (2003). A relational view of information seeking and learning in social networks. *Management Science 49*(4), 432–445.

Calvo-Armengol, A. and M. Jackson (2004). The effects of social networks on employment and inequality. *American Economic Review 94*(3), 426–454.

Calvo-Armengol, A. and M. Jackson (2007). Networks in labor markets: Wage and employment dynamics and inequality. *Journal of Economic Theory 132*(1), 27–46.

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

Chandrasekhar, A. and R. Lewis (2011). Econometrics of sampled networks. Mimeo, MIT Department of Economics.

Chidambaran, N., S. Kedia, and N. Prabhala (2012). CEO-Director connections and corporate fraud. SSRN Working Paper No. 2023030.

Ductor, L., M. Fafchamps, S. Goyal, and M. van der Leij (2013). Social networks and research output. *Review of Economics and Statistics forthcoming*.

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*(8), 1455–1475.

Gonul, F. and K. Srinivasan (1993). Modeling multiple sources of heterogeneity in multinomial logit models: Methodological and managerial issues. *Marketing Science 12*(3), 213–229.

Gonul, F. and K. Srinivasan (1996). Estimating the impact of consumer expectations of coupons on purchase behavior: A dynamic structural model. *Marketing Science 15*(3), 262–279.

Gulati, R. (1999). Network location and learning: The influence of network resources and firm capabilities on alliance formation. *Strategic Management Journal 20*(5), 397–420.

Guryan, J., K. Kroft, and M. Notowidigdo (2009). Peer effects in the workplace: Evidence from random groupings in professional golf tournaments. *American Economic Journal: Applied Economics 1*(4), 34–68.

Heckman, J. (1981). Structural analysis of discrete data with econometric applications. In C. Manski and D. McFadden (Eds.), *Statistical models for discrete panel data*, pp. 179–195. Cambridge MA: MIT Press.

Ibarra, H. (1997). Paving an alternative route: Gender differences in managerial networks. *Social Psychology Quarterly 60*(1), 91–102.

Jackson, M. and B. Rogers (2005). The economics of small worlds. *Journal of the European Economic Association 3*(2-3), 617–627.

Jackson, M. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*(1), 44–74.

Jain, D., N. Vilcassim, and P. Chintagunta (1994). A random-coefficients logit brand-choice model applied to panel data. *Journal of Business and Economic Statistics 12*(3), 317–328.

Jain, T. and M. Kapoor (2012). The impact of study group and residential peers on academic performance. Indian School of Business Working Paper No. ISB-WP/103/2012.

Lerner, J. and U. Malmendier (2012). With a little help from my (random) friends: Success and failure in post-business school entrepreneurship. *Review of Financial Studies forthcoming*.

Maitra, P. and S. Mani (2012). Learning and earning: Evidence from a randomized evaluation in India. Fordham University Discussion Paper Series DP2013-02.

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

Mas, A. and E. Moretti (2009). Peers at work. *American Economic Review 99*(1), 112–145.

Nanda, R. and J. Sørensen (2010). Workplace peers and entrepreneurship. *Management Science 56*(7), 1116–1126.

Rossi, P., G. Allenby, and R. McCulloch (2005). *Bayesian Statistics and Marketing*. New York NY: John Wiley and Sons.

Rossi, P., R. McCulloch, and G. Allenby (1996). The value of purchase history data in target marketing. *Marketing Science 15*(4), 321–340.

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

Shalizi, C. and A. Thomas (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research 40*(2), 211–239.

Shue, K. (2013). Executive networks and firm policies: Evidence from the random assignment of MBA peers. *Review of Financial Studies forthcoming*.

Stinebrickner, R. and T. Stinebrickner (2006). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics 90*(8-9), 1435–1454.

A GPA as a measure of productivity

To verify that students are motivated to maximize their grades, we specify the following model and examine the correlation between core term grades and post-program earnings.

$$y_{ijt} = \gamma_0 + \gamma_1 core\_GPA_{ijt} + \gamma_2 elective\_GPA_{ijt} + \gamma_3 GMAT_{ijt} + \gamma_4 X_{ijt} + year_t + \epsilon_{ijt} \tag{7}$$

In this regression, y_{ijt} is the value of the job offer reported by a student after on-campus interviews. Although a student might receive multiple job offers, we use the salary associated with the accepted job. The coefficient of interest is γ_1, which represents the impact of a student's cumulative GPA at the end of core terms on the salary, after controlling for elective terms GPA. The coefficients in γ_4 represent the impact of other individual factors, such as whether the student has a master's degree, the number of years of experience, marital status, age, female, last salary before business school and citizenship status. We also include indicator variables for students who attend either Delhi University or Indian Institute of Technology, since the largest fraction of students attended these universities for undergraduate studies. Finally, cohort fixed-effects account for cohort-wide changes in academic policies as well as variations in economic conditions which might influence earnings. Note that the coefficients of this model cannot be interpreted as causal estimates since we cannot rule out the impact of unobserved factors that impact both core terms GPA and earnings.

Table 11 reports the results of the estimation exercise and shows that core terms GPA (on a continuous scale of 0 to 4) is very strongly correlated with salary. Increasing GPA by one point is associated with an increase of Rs. 331329.5 in reported salary ($p < 0.01$). This result is not surprising. In contrast to several major business schools that follow grade non-disclosure policies, ISB permits students to report their GPAs to potential employers who use this information to screen interview candidates.[17]

Although elective terms GPA also influences earnings, the impact is smaller. Not only is the associated coefficient (167964.5) less than half the size of the coefficient associated with core term grades, but also statistically insignificant. Other characteristics that significantly influence reported

[17]For example, see the section on education from a sample resume in Figure 3. Anecdotal evidence suggests that consulting firms, which hire approximately one third of students, screen on the basis of GPA only, and often ignore other factors such as past work experience or specialization.

salaries are years of work experience and citizenship. However, since a student cannot change any of these factors while at ISB, the results suggest that students are strongly motivated to maximize their GPA in the core terms.

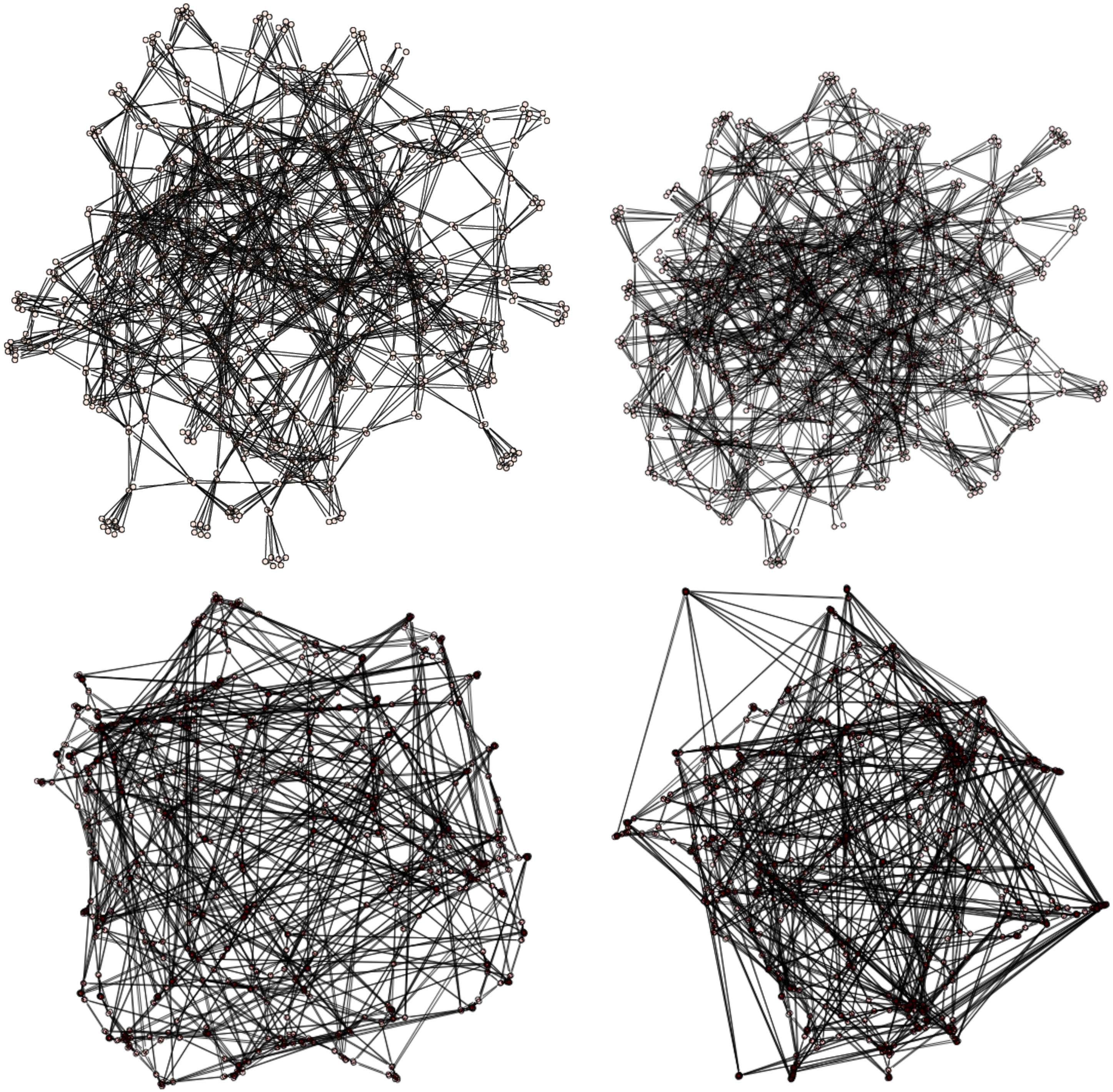

Figure 2: **Networks for 2007-08, 2008-09, 2009-10 and 2010-11 cohorts**

Table 1: **ISB compared to international business schools**

	GMAT (Mean)	Years of work experience	Female (Fraction)	Class size
Stanford GSB	730	4.1	34%	401
Harvard Business School	724	4.0	39%	901
Wharton (UPenn)	720	6	36%	823
Kellogg (Northwestern)	715	5	35%	475
Booth (UChicago)	715	4.6	35%	1177
IIM Ahmedabad PGPX	713	10	7%	86
Indian School of Business	**712**	**4.9**	**28%**	**560**
MIT Sloan	710	5	35%	396
INSEAD	703	6	33%	988
Darden (University of Virginia)	701	4.7	29%	328
Fuqua (Duke)	698	5.0	37%	887
London Business School	694	5.6	25%	319

Note: This table reports summary characteristics of students enroled at select international business schools. The data is for the Class of 2011 for the full time MBA programs (or equivalent) for all schools. Source: School websites and `http://www.businessweek.com`.

Table 2: **Summary statistics**

Variable	**Observations**	**Mean**	**Standard Deviation**
Years of experience	1987	4.9	2.2
Pre-program earnings (Rs. ’00000)	1835	10.0	12.0
GMAT	1987	709.0	40.2
Masters degree	1987	0.19	0.40
IIT	1987	0.14	0.35
Delhi University	1987	0.15	0.36
Single	1987	0.73	0.44
Age	1987	28.7	2.8
Female	1987	0.26	0.44
Citizen of India	1987	0.96	0.20
Post-program earnings (Rs. ’00000)	1755	14.01	8.69
Degree centrality	1976	5.86	1.34
Closeness centrality	1976	0.00041	0.00007

Notes: This table shows the summary statistics for the main dataset. Raw measures (not z-scores) of degree and closeness centrality are reported. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. Source: ISB administrative records.

Table 3: **Randomization check**

	Female		**Male**	
	Coefficient	**Standard Error**	**Coefficient**	**Standard Error**
I. GMAT score				
Study group mean score	0.084	(0.099)	0.023	(0.054)
Quad mean score	-0.064	(0.081)	0.022	(0.051)
Degree centrality	-7.136	(4.399)	1.787	(2.715)
Closeness centrality	1.474	(7.316)	-5.308	(4.178)
Observations	434		1069	
R-squared	0.048		0.040	
II. Pre-program earnings				
Study group mean earnings	0.077	(0.140)	0.007	(0.046)
Quad mean earnings	0.023	(0.085)	-0.053	(0.054)
Degree centrality	0.183	(1.920)	-0.791	(0.660)
Closeness centrality	-0.209	(3.264)	1.411	(1.053)
Observations	406		986	
R-squared	0.006		0.064	

Notes: This table examines the correlation between a student's GMAT score and last earnings before business school, and the mean characteristics of the study groups and roommates, including network measures. The results are reported separately for sub-samples of men and women. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. OLS specifications include year fixed effects. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

Table 4: **Allocation to single apartments**

Dependent variable: Residence in a single apartment				
	OLS		**Probit**	
	Coefficient	**Standard Error**	**Marginal effect**	**Standard Error**
GMAT	0.000	(0.000)	0.000	(0.000)
Years of experience	0.004	(0.007)	0.002	(0.009)
Pre-program earnings	0.001	(0.001)	0.001	(0.001)
Masters degree	0.004	(0.022)	0.007	(0.027)
Age	0.007	(0.006)	0.010	(0.007)
Single	**-0.556*** **	(0.022)	**-0.564*** **	(0.029)
Female	**-0.095*** **	(0.019)	**-0.114*** **	(0.020)
Citizen of India	**-0.222*** **	(0.041)	**-0.310*** **	(0.070)
Observations	1985		1985	
R-squared	0.401		0.349	

Notes: This table examines the characteristics of students who were allocated single apartments (without roommates) at ISB. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. The dependent variable is 1 if the student lived in a single apartment and 0 if the student lived in a shared quad while enrolled at ISB. Both the OLS and probit regression specifications include year fixed-effects. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

Table 5: **Impact of networks on core terms GPA**

Dependent variable: Core terms GPA					
	I	**II**	**III**	**IV**	**V**
$Experience_i$	**-0.053*****	**-0.053*****	**-0.054*****	**-0.053*****	**-0.053*****
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
$GMAT_i$	**0.103*****	**0.102*****	**0.107*****	**0.107*****	**0.107*****
	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)
$Pre - earnings_i$	**0.025*****	**0.025*****	**0.024*****	**0.024*****	**0.023*****
	(0.00637)	(0.00637)	(0.00633)	(0.00633)	(0.00631)
$Degree_i$		-0.004	-0.004	0.005	0.003
		(0.011)	(0.010)	(0.022)	(0.022)
$Closeness_i$		**-0.050****	**-0.051****	-0.081	-0.094
		(0.021)	(0.021)	(0.095)	(0.094)
$GMAT_i * Degree_i$			0.006	0.005	0.005
			(0.007)	(0.006)	(0.007)
$GMAT_i * Closeness_i$			**-0.033*****	**-0.033*****	**-0.033*****
			(0.007)	(0.007)	(0.007)
$Mean(Degree_{-i})$				-0.016	-0.018
				(0.014)	(0.014)
$Mean(Closeness_{-i})$				0.057	0.075
				(0.107)	(0.107)
$Max(GMAT_{-i})$					**0.025*****
					(0.007)
Observations	1835	1824	1824	1824	1824
R-squared	0.287	0.292	0.302	0.303	0.308

Notes: This table examines the impact of network measures on individual GPA. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. The dependent variable is each student's GPA from 16 required courses in the one-year program. The table reports coefficients obtained from OLS estimation of equation (3). The specification includes variables for gender, marital status, citizenship, masters degree, undergraduate institution, year fixed effects. Standard error in parentheses. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

Table 6: **Heterogeneous impact of networks on core terms GPA**

	Base	**High GMAT**	**Low GMAT**	**Female**	**Male**
$Experience_i$	**-0.053*****	**-0.046*****	**-0.060*****	-0.022	**-0.062*****
	(0.008)	(0.013)	(0.009)	(0.016)	(0.009)
$GMAT_i$	**0.107*****	**0.142*****	**0.095*****	**0.114*****	**0.107*****
	(0.007)	(0.022)	(0.011)	(0.012)	(0.008)
$Pre-earnings_i$	**0.023*****	**0.051*****	0.013	0.004	**0.045*****
	(0.006)	(0.012)	(0.007)	(0.008)	(0.009)
$Degree_i$	0.003	0.012	0.041	0.064	-0.025
	(0.022)	(0.036)	(0.031)	(0.043)	(0.026)
$Closeness_i$	-0.094	0.0313	-0.216	-0.336	-0.001
	(0.094)	(0.142)	(0.128)	(0.185)	(0.111)
$GMAT_i * Degree_i$	0.005	**-0.052****	**0.022****	-0.011	0.013
	(0.006)	(0.023)	(0.010)	(0.014)	(0.007)
$GMAT_i * Closeness_i$	**-0.033*****	0.004	**-0.067*****	**-0.049*****	**-0.026****
	(0.007)	(0.023)	(0.011)	(0.012)	(0.008)
$Mean(Degree_{-i})$	-0.018	-0.004	-0.029	-0.042	-0.010
	(0.014)	(0.020)	(0.019)	(0.029)	(0.016)
$Mean(Closeness_{-i})$	0.075	-0.054	0.151	0.281	-0.009
	(0.107)	(0.158)	(0.145)	(0.213)	(0.124)
$Max(GMAT_{-i})$	**0.025*****	**0.024***	**0.025****	**0.034****	**0.023****
	(0.007)	(0.010)	(0.010)	(0.013)	(0.008)
Observations	1824	862	962	479	1345
R-squared	0.308	0.231	0.257	0.296	0.324

Notes: This table reports the impact of heterogeneity in gender and ability on the effect of network measures on individual GPA. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. The dependent variable is each student's GPA from 16 required courses in the one-year program. The table reports coefficients obtained from OLS estimation of equation (3). The specification includes variables for gender, marital status, citizenship, masters degree, undergraduate institution, year fixed effects. Standard error in parentheses. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

Table 7: **Bayesian coefficients**

	Coefficient	**Standard error**
$Experience$	**-0.05*****	(0.001)
$GMAT$	**0.124*****	(0.001)
$Pre-earnings$	**0.071*****	(0.002)
$Degree$	-0.0001	(0.003)
$Closeness$	**-0.073*****	(0.014)
$GMAT * Degree$	0.000	(0.001)
$GMAT * Closeness$	**-0.019*****	(0.001)
$Mean(Degree_{-i})$	0.002	(0.002)
$Mean(Closeness_{-i})$	-0.012	(0.016)
$Max(GMAT_{-i})$	**0.023*****	(0.001)

Notes: This table reports Bayesian coefficients obtained from estimation of equation (4), on the effect of network measures on individual GPA. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. The dependent variable is each student's GPA from 16 required courses in the one-year program. The specification includes variables for gender, marital status, citizenship, masters degree, undergraduate institution, year fixed effects. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

Table 8: **Heterogeneous impact of networks on core terms GPA (Bayesian analysis)**

Dependent variable: Core terms GPA						
	Intercept		**High GMAT**		**Female**	
$Experience_i$	**-0.042*****	(0.002)	0.004	(0.003)	**-0.02*****	(0.003)
$GMAT_i$	**0.138*****	(0.002)	**-0.028*****	(0.003)	**-0.016**	(0.002)
$Pre-earnings_i$	**0.064*****	(0.003)	**-0.011*****	(0.004)	**0.021*****	(0.003)
$Degree_i$	**0.083*****	(0.004)	**-0.234*****	(0.010)	**-0.051*****	(0.007)
$Closeness_i$	**-0.472*****	(0.017)	**0.963*****	(0.044)	**0.326*****	(0.028)
$GMAT_i * Degree_i$	**-0.015*****	(0.002)	**0.02*****	(0.003)	**0.023*****	(0.003)
$GMAT_i * Closeness_i$	**-0.016*****	(0.002)	**0.031*****	(0.003)	**-0.022*****	(0.003)
$Mean(Degree_{-i})$	**-0.023*****	(0.003)	**0.093*****	(0.005)	0.003	(0.004)
$Mean(Closeness_{-i})$	**0.361*****	(0.019)	**-0.982*****	(0.048)	**-0.261*****	(0.031)
$Max(GMAT_{-i})$	**0.019*****	(0.002)	**-0.006****	(0.003)	**0.012*****	(0.002)

Notes: This table reports Bayesian coefficients obtained from estimation of equation (4), showing the impact of heterogeneity in gender and ability on the effect of network measures on individual GPA. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. The dependent variable is each student's GPA from 16 required courses in the one-year program. The specification includes variables for gender, marital status, citizenship, masters degree, undergraduate institution, year fixed effects. Standard errors in parentheses. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

Table 9: **Impact of network measures by groupwork component in course**

	All core courses GPA	High Groupwork	Medium Groupwork	Low Groupwork
$GMAT_i$	**0.107*****	**0.082*****	**0.109*****	**0.145*****
	(0.007)	(0.006)	(0.008)	(0.009)
$Degree_i$	0.003	0.016	0.005	-0.014
	(0.022)	(0.021)	(0.026)	(0.030)
$Closeness_i$	-0.094	-0.134	-0.134	-0.019
	(0.095)	(0.089)	(0.113)	(0.128)
$GMAT_i * Degree_i$	0.005	0.0004	0.0103	0.007
	(0.006)	(0.006)	(0.008)	(0.009)
$GMAT_i * Closeness_i$	**-0.033*****	**-0.023*****	**-0.040*****	**-0.041*****
	(0.007)	(0.006)	(0.008)	(0.009)
$Mean(Degree_{-i})$	-0.018	-0.015	-0.018	-0.017
	(0.014)	(0.013)	(0.017)	(0.019)
$Mean(Closeness_{-i})$	0.075	0.101	0.112	0.016
	(0.107)	(0.100)	(0.127)	(0.144)
$Max(GMAT_{-i})$	**0.025*****	**0.025*****	**0.024****	**0.024****
	(0.007)	(0.006)	(0.008)	(0.009)
Observations	1824	1824	1824	1824
R-squared	0.308	0.256	0.244	0.300

Notes: This table examines the impact of network measures on GPA calculated separately for courses with high, medium and low group work component of the final grade. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. Table reports coefficients obtained from OLS estimation of equation (3) performed for the GPA averaged over all courses, courses with high group work component of final grade (35 to 50%), courses with medium group work component of final grade (20 to 30%) and courses with low group work component of final grade (0 to 15%). Regression specification includes year fixed effects, as well as variables for years of experience, last salary, masters degree, undergraduate institution, gender, marital status, citizenship and age in each category. Standard errors in parentheses. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

Table 10: **Impact of network measures over terms**

	Core terms GPA	**Term 1 GPA**	**Term 2 GPA**	**Term 3 GPA**	**Term 4 GPA**
$GMAT_i$	**0.107*****	**0.131*****	**0.094*****	**0.087*****	**0.116*****
	(0.007)	(0.009)	(0.007)	(0.008)	(0.008)
$Degree_i$	0.003	-0.027	0.026	-0.004	0.016
	(0.022)	(0.030)	(0.024)	(0.030)	(0.026)
$Closeness_i$	-0.094	0.018	-0.179	-0.075	-0.141
	(0.094)	(0.126)	(0.102)	(0.113)	(0.109)
$GMAT_i * Degree_i$	0.005	0.010	0.0001	0.002	0.008
	(0.007)	(0.009)	(0.007)	(0.008)	(0.007)
$GMAT_i * Closeness_i$	**-0.033*****	**-0.034*****	**-0.028*****	**-0.032*****	**-0.037*****
	(0.007)	(0.009)	(0.007)	(0.008)	(0.008)
$Mean(Degree_{-i})$	-0.018	-0.003	**-0.036***	-0.014	-0.019
	(0.014)	(0.019)	(0.015)	(0.017)	(0.016)
$Mean(Closeness_{-i})$	0.075	-0.076	0.184	0.062	0.133
	(0.107)	(0.142)	(0.115)	(0.128)	(0.123)
$Max(GMAT_{-i})$	**0.022*****	**0.028****	**0.026*****	**0.024****	**0.023****
	(0.007)	(0.009)	(0.007)	(0.008)	(0.008)
Observations	1824	1824	1824	1824	1824
R-squared	0.308	0.279	0.230	0.218	0.243

Notes: This table examines the impact of network measures on GPA calculated separately for courses in terms 1, 2, 3 and 4, and on earnings obtained after business school. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. Table reports coefficients obtained from OLS estimation of the main specification. Regression specification includes year fixed effects, as well as variables for years of experience, last salary, masters degree, undergraduate institution, gender, marital status, citizenship and age in each category. Standard errors in parentheses. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

Table 11: **Determinants of earnings**

Dependent variable: Annual earnings from on-campus placement		
	Coefficient	**Standard error**
Core terms GPA	**286172.4****	(95318.6)
Elective terms GPA	174981.0	(94324.9)
GMAT	855.8	(557.4)
Full time experience	**48358.8****	(17862.4)
Pre-program earnings	-1516.7	(1717.2)
Masters	-12199.6	(54808.7)
IIT	112906.5	(61006.4)
Delhi University	-19779.9	(56070.7)
Citizen of India	**355984.7****	(109352.3)
Female	-12292.0	(47674.4)
Single	-633.2	(55068.7)
Age	7032.5	(15513.2)
Observations	1605	
R-squared	0.157	

Notes: This table reports the correlation of individual characteristics with the earning associated with the accepted job obtained through the on-campus placement office. Each observation is a student, and we pool the sample over all class years from 2007-08, 2008-09, 2009-10 and 2010-11 class years. The table reports coefficients obtained from OLS estimation of equation (7). The regression controls for year fixed effects. *** implies significance at the 0.01 level, ** 0.05, * 0.10. Source: ISB administrative records.

EDUCATION

Indian School of Business **April 2009–to date**

Post Graduate Program in Management (Major- Finance & Marketing)

- **CGPA: 3.59/4.0 (Top 16%)**; GMAT: 760/800 (99 percentile)
- Developing the growth strategy for Staples-Future Group JV to achieve 40% growth in revenues through the retail channel
  - Designed the Customer Acquisition and Retention programme to be implemented across stores nationwide
  - Carried out extensive market and competitor analysis to come up with a new sales and operations model
- Undertook a consulting engagement with ISB Operations and Marketing teams to redesign the ISB Merchandize Store
  - Project being implemented for Solstice'09- annual ISB alumni meet
- Working on a Research Project on the transformation of State Bank of India through Business Process Re-engineering

Indian Institute of Technology, Kanpur **May 2002-May 2006**

Bachelor of Technology (B.Tech), Electrical Engineering

- **CGPA: 8.4/10;** Among the **top 0.1%** in IIT Joint Entrance Exam (All India Rank 255)
- Summer Training at ITC Ltd.: Designed a benchmark cigarette factory based on Lean manufacturing principles (Capex proposal of $23 million). Project adjudged among the top 3 projects (from 20+ projects) in the division

Delhi Public School, R.K. Puram **April 2001- March 2002**

- Secured **93.2%** in CBSE, AISSCE; **Ranked 7/889** in school; Awarded National CBSE Merit Scholarship and Gold Medal

Figure 3: **Sample resume**