Does Who You Know Matter? Unraveling the Influence of Student Networks on Academic Performance

Tarun Jain
Indian School of Business

Nishtha Langer¹

“It is not what you know but who you know that makes a difference” – Anon.

Motivation and Background

Conventional wisdom proclaims that networks are critical to human capital formation and therefore should result in higher productivity. Although extant research finds that human capital investments such as training improve productivity (Bapna et al. 2012), it is unclear whether peer networks also impact productivity, and if so, the magnitude, mechanisms and implications of these effects. In this paper, we examine the effect of student networks on academic performance. Although a rich literature examines peer effects in the context of education (e.g., Sacerdote 2001, Zimmerman 2003, Carrell, Fullerton, and West 2009), the focus has been on dyadic relationships in relatively small groups. Researchers have also been unable to quantify the impact of complete network structures, primarily because of data limitations. Mapping of peers beyond the immediate network, the nature of these relationships, and their impact on individual productivity has been under researched in the literature.

We address this gap in the literature by bringing in new measures of network structure to the human capital literature. In particular, we examine the following research questions in an academic setting: (1) What should the network map for a student look like? (2) What are the relevant network measures, and what are their economic implications? (3) What is the effect of an individual’s network measures on productivity?

In addition, using data from a business school where students are assigned exogenously to interlinked networks of different sizes, we address the critical empirical issue of identification of peer or network effects separately from self-selection into favorable networks due to homophily or correlated effects due to shared environment within networks (Manski 1993, McPherson et al. 2001).

We find that increasing centrality within the network is detrimental to student performance. In addition, connections to highly connected individuals have no significant effect on performance. However, connections to individuals who are well connected and who have high ability are associated with high productivity.

To summarize, we contribute to the literature by studying a complete and exogenous network structure, which helps us in identification of the peer effects. Secondly, we focus on network measures such degree and closeness centrality that helps us measure the peer effects beyond that of immediate network. Thirdly, our study examines outcome measures that are well

¹ Corresponding author, email: nishtha_langer@isb.edu.
defined and objective. Finally, unlike Zimmerman et al., our study examines productivity in a setting where networks should be influential, extending the work of Jain and Kapoor (2012). We also contribute to the literature on familiarity (Espinosa et al. 2007), wherein we not only measure the effect of team level dyads but also the effect of the network structure on individual performance.

**Theoretical Underpinnings and Empirical Challenges**

Network theory (Brass 2002) posits that individuals who are more centrally located or have many ties are better able to aggregate information. Indeed, as Alatas et al. (2012) find, nodes that are better connected are also better informed. Networks also facilitate communication, whereby information and knowledge can be exchanged across individuals who form the network (e.g., Reagans and Zuckerman 2001). Thus, we should expect that individuals who are centrally located, due to access to pertinent and useful information, are likely more productive.

However, being centrally located within a network also places certain constraints on each individual. While the individual may benefit from access to information and knowledge, information search is costly. First, there are transactions costs associated with network interactions. For instance, an individual will interact with the network to search for pertinent information (Simon 1978, Uzzi 1997). Second, a more central individual may be connecting various parts of the network, and hence this further increases her opportunity costs. Finally, the subsequent to such network interactions, the individuals also need to process information and prioritize such information so that it can be productive. Such paradox of embeddedness has been well documented in the sociology literature (Uzzi 1996, 1997). Thus, *we argue that centrality is negatively associated with academic performance.*

We also examine how the relational structure of networks affects academic performance. Borgatti and Cross (2003) suggest that both network centrality and expertise (knowing) will influence an individual’s information seeking behavior. Therefore, we hypothesize that *being connected to individuals who are well connected is positively associated with academic performance.* Furthermore, we argue that *connecting to an able and more connected individual is positively associated with academic performance.*

A number of empirical challenges are associated with research on network effects. First, rarely do researchers have information about the complete network structure. For instance, nodes sampled within the network may be systematically different from unsampled nodes. Missing nodes will also lead to underestimating the centrality of sampled nodes within the network. Therefore, sampling would bias estimates of network effects. Second, even when complete information about the network structure is available, the network effects may be confounded with other endogenous effects (Manski 1993). Networks tend to form between people who believe the association will be beneficial, and hence we have to contend with the issue of self-selection. Third, networks often manifest homophily effects (see Aral and Walker 2011 for an excellent discussion). People close to each other in a network may share similar attributes; they associate with each other due to mutually shared interests, and may be influenced by shared
environmental characteristics. Thus, assortative traits, or homophily, may be misattributed as network effects. All of these factors make it challenging to identify the effect of network characteristics on productivity.

**Research Setting and Data**

Research on peer effects in an academic setting, wherein members are randomly (and exogenously) assigned to the network, addresses the concerns outlined above (Sacerdote 2001). Our data come from the flagship post-graduate business program at the Indian School of Business (ISB). ISB is a large, independent provider of post-graduate management education established in 2001 with a one-year, full-time residential diploma program. Since 2009, the Financial Times newspaper has ranked the program among top 20 MBA programs in the world. An application to ISB consists of GMAT scores, essays, letters of recommendation, undergraduate and graduate transcripts, and an interview. The GMAT scores, the prior work experience, and gender diversity of the applicants are comparable to those of leading business schools in the world, ISB is thus arguably similar to a number of major international business schools on observable characteristics. The students take a common “core” of 16 non-elective classes in the first four terms; and choose from various elective courses that allow them to complete their major requirements in the remaining four terms.

We obtained detailed records on GPA as well as assignment of students to study groups and residential facilities for four years from 2007-08 to 2010-11. One advantage of selecting this period was the absence of significant changes in the curriculum or administrative policies during this time. We supplement this 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, earning and functional role), and demographic backgrounds (year of birth, gender, marital status, and citizenship). This combined dataset offers a number of features that makes it attractive for analyzing peer effects on the academic performance of business school students. First, the administrative source of the data allows us to map the entire set of formal and informal peers for each student, and avoid potentially biased estimates inherent in partial network sampling (Chandrasekhar and Lewis 2011). We used the combined group information to construct the network, as shown in figure 1A; the entire network for 2007-08 is illustrated in Figure 1B.

---

2 An administrator assigns four to five students to the core study group (for the duration of the core terms), and four members to residential group (for the program duration), wherein they share a quad. This assignment is random in that the administrator does not assign on the basis of ability. However, gender is the primary criterion for such assignment, a study group should contain at least two or no women, and the residential group should be of the same gender. The residential groups are assigned such that there is no overlap with the study group. A student may choose instead to stay in a studio apartment. However, this choice is mostly exercised by students who cohabitate with family. T-tests of means for core GPA and GMAT for those who stay in quads with their residential groups versus those who stay in studio apartments indicated no significant differences, suggesting that ability does not dictate this choice.

3 Since all administrative records are mandated to be complete and truthful, self-reporting bias, measurement error, and missing data do not threaten our analysis. We also find that attrition is negligible in the one year PGP program, and student cohorts do not overlap. Hence, non-random attrition from the sample and serial correlation due to overlapping peers across years are not a significant concern.
We compute network measures such as degree centrality and closeness centrality for our analysis. Note, however, that we rely on the random assignment of students to these different groups, rather on any measure of ability, to identify our model. To this end, we regress these network variables on individual GMAT scores, including year dummies as our control variables. Our results lend credence to our belief that the administrative randomization process led to the network structure where ability was uncorrelated.

**Analysis and Results**

Our variable of interest in terms of academic performance is the core term grade point average (GPA),\(^5\) for multiple reasons. First, all the students take the same set of core courses. Second, the elective group study groups are entirely self-selected, and hence not useful for the purpose of this study. Third, it offers us an objective way of assessing a student’s performance and students are strongly motivated to maximize the core GPA.\(^6\) In particular, we specify and estimate the following model:

\[
\text{Core GPA}_{ijt} = \beta_0 + \beta_1 X_{ijt} + \beta_2 \text{GMAT}_{ijt} + \beta_3 \text{Degree}_{ijt} + \beta_4 \text{Closeness}_{ijt} + \beta_5 \text{YearFE} \\
+ \beta_6 \text{Degree}_{ijt} + \beta_7 \text{Closeness}_{ijt} + \beta_8 \text{GMAT}_{ijt} + \epsilon_{ijt}
\]

Here, CORE_GPA is the GPA for the core terms for student ‘i’ in group ‘j’ in cohort ‘t.’ \(X_{ijt}\) is a vector of individual characteristics such as full time experience and age in years, and the last salary prior to joining the program. These variables help capture student maturity, experience with solving business problems, and success in the corporate world, and may be indicators of an individual’s drive and motivation. We also include gender, marital status, and citizenship. We include tier 1 school dummies (IIT, Delhi U) for individual ‘i’ and year fixed effects to control for

\(^4\) Since gender is a criterion for the group assignment, we report our results separately for men and women. For women (men), the t-stats for a regression with \(H_0\) that degree centrality and closeness centrality are uncorrelated with GMAT are -0.24 (1.37) and -0.30 (0.61) respectively.

\(^5\) Instructors at ISB award course grades on a four-point scale. The highest grades 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. Manifestly, a student’s objective is to earn the maximum score possible, regardless of the relative performance of the other members of the study or residential groups. We also focus on core term GPA rather than the overall GPA since the choice of electives may confound our findings.

\(^6\) It is possible that the end goal of a student from this program is to maximize placement offers rather than the core GPA. However, we note that unlike other top rated business schools with grade non-disclosure policies, ISB graduates are allowed to share their GPA with prospective employers. We regressed placement offers with core GPA, while controlling for demographic and other controls. We find that a one point increase in the GPA is associated with an increase of Rs. 483,073 (about $8,950). While other factors such as experience are also significant determinants of these offers, none of these factors can be changed by the student while at ISB. The results are available from the authors by request.
various observed and unobserved factors. We include an individual’s GMAT as a proxy for her academic ability. Variables degree and closeness are the network variables, the subscript ‘i’ indicates that these refer to the individual, and subscripts ‘-i’ refer to the max within the group ‘j’ excluding ‘i.’ The variable GMAT$_{ijt}$ is the GMAT of group member with the highest degree centrality in the group ‘j.’

Table 1 reports the results of this estimation. We estimate a baseline model and subsequently add the network variables of interest. Not surprisingly, an individual’s own GMAT score is highly correlated with the GPA. Last salary, which may proxy for an individual’s drive and motivation, positively associated with GPA. This variable also likely captures job and industry-specific ability that is different from an academic ability captured by the GMAT score. Married students on an average have 0.101 grade point advantage over unmarried students, consistent with empirical research that observes that married workers have higher earnings than unmarried ones (Lundberg and Ross 2000). We find that women, on the other hand, seem worse off than men, perhaps for reasons outside the scope of this paper.

Our results show that degree centrality is negatively associated with GPA, while closeness centrality is not significant. The former speaks to the opportunity costs of being too central and embedded in the network, whereas the latter results indicate that local rather than network wide peer effects may be more influential in predicting academic performance. Surprisingly, we find that access to well-connected individuals in the group has no effect on GPA. In contrast, the GMAT of the most well-connected individual in the group is not only statistically significant; it has almost a quarter of an impact of one’s own GMAT score on GPA.

We also examine whether individuals are differently affected by the network structure. For instance, students with low ability (low GMAT) are more likely to depend on the network for...
information exchange and learning, but may also be less efficient in allocating their time and have higher opportunity costs from network interactions. Our preliminary sub-sample analysis shows (not included here, but available on request) that for low ability students (as measured by GMAT), higher degree centrality is associated with worse performance, compared to the high ability students (for whom the effect is not significant). Further, access to high ability students in the group who are also well-connected leads to better outcomes for these students (again, for high ability students, this is not significant).

These results encourage us to pursue a hierarchical Bayesian specification that allows us to exploit individual level heterogeneity (Rossi et al. 1996, 2005). Furthermore, it is likely that groups learn over a period of time, and hence the effect of information flow in the network and the transactions costs associated with networks also change over a period of time. To address this, we plan on using data for the first four core terms to constitute a panel, and specifying a dynamic panel model (Arellano and Bover 1995; Blundell and Bond 1998) to help with identification.

**Discussion and Conclusion**

In this paper, we examine the effect of peer level networks on academic performance, using data from an elite business school. Our results indicate that while being connected to individuals with high ability and high degree centrality is beneficial, there are costs associated with being too central to the network. In essence, *it is not who you know, but what they know.*

These results have important implications not only for performance in other educational institutes but also for human capital formation and workplace productivity. We used GPA as a measure of productivity. We find that after controlling for various demographic and other variables, it is a robust predictor of short term earnings (e.g., placement offers), and we expect it to affect long term earnings as well. We expect that this research to have both theoretical and practical significance. For instance, our research underlines the importance of being connected to knowledgeable nodes, but also highlights the dichotomy of being too connected. An important implication of this research is how to disseminate information and knowledge within the network, but at the same time lowering the consequent transaction costs.

Like all empirical research, our research has certain limitations. For instance, it is possible that the student networks have unobservable characteristics that may also influence productivity. However, we believe that learning and consequently productivity outcomes are unlikely to be influenced by such network attributes, given that network assignment is largely random and exogenous. It is also possible that factors such as motivation and drive, which are not captured by our model, are likely to influence productivity. Finally, these students may be part of other informal networks that are positively associated with GPA. To that end, we intend to use (undergraduate) alma mater and/or hometown location as instruments that inform our network variables. We look forward to presenting the full set of results at WISE 2012!

**References** are available from the authors.