Research Engineer Network: A Network Analysis of Graduate Student Relationships

Dr. Chrysafis Vogiatzis, University of Illinois at Urbana-Champaign

Dr. Chrysafis Vogiatzis is a teaching assistant professor for the Department of Industrial and Enterprise Systems Engineering at the University of Illinois at Urbana-Champaign. Prior to that, Dr. Vogiatzis was an assistant professor at North Carolina Agricultural and Technical State University. His current research interests lie in network optimization and combinatorial optimization, along with their vast applications in modern socio-technical and biological systems. He is serving as the faculty advisor of the Institute of Industrial and Systems Engineers, and was awarded the 2019 Faculty Advisor award for the North-Central region of IISE.

Dr. Stephanie Marie Teixeira-Poit, North Carolina A&T State University

Stephanie M. Teixeira-Poit, PhD, Assistant Professor of Sociology at North Carolina A&T State University, leads large-scale, mixed-methods projects that seek to address disparities through complex intervention implementation and evaluation. Dr. Teixeira-Poit has three primary research streams. First, she implements and evaluates interventions to address workforce shortages and improve the capacity of the workforce. Second, she leads health services studies that examine the impact of developing systems of care and transforming practices on health care access and utilization, delivery and quality of care, and health outcomes. Third, she assesses the effect of social determinants of health on access to care and patient outcomes. She evaluates the effectiveness of interventions designed to attenuate the effect of social determinants on patient outcomes. She has 15 years of experience leading research teams; designing and implementing research and evaluation; developing protocols for surveys, interviews, and focus groups; collecting and analyzing qualitative data, and programming advanced statistical analyses of quantitative data using Stata. She has served as principal investigator or task leader on research studies funded by the U.S. Centers for Disease Control and Prevention (CDC), Centers for Medicare and Medicaid Services (CMS), Agency for Healthcare Research and Quality (AHRQ), Health Resources and Services Administration (HRSA), National Cancer Institute (NCI), National Multiple Sclerosis Society (NMSS), and National Science Foundation. Before joining the University, Dr. Teixeira-Poit worked as a researcher for 7 years at RTI International and 5 years at the Center for Urban Affairs and Community Services.

Dr. Tobin N. Walton, North Carolina A&T State University

My research is focused on developing interdisciplinary theoretical frameworks and methodological designs capable of modeling the social and psychological drivers of behavior, decision-making, and information processing across multiple domains (e.g., STEM education, food security, the environment).

Dr. Grace Gowdy, North Carolina A&T State University

Dr. Gowdy is an Assistant Professor at North Carolina A&T's Department of Social Work & Sociology. Dr. Gowdy currently works on multiple studies examining how formal and informal mentoring relationships can support educational outcomes for historically-underrepresented students.

Dr. Bala Ram P.E., North Carolina A&T State University

Dr. Bala Ram serves as a Professor in Industrial Systems Engineering and the Associate Dean for Academic Affairs for the College of Engineering at NC A&T State University. As a faculty member, he played a significant role in the implementation of a PhD in Industrial & Systems Engineering. Dr. Ram served as the PI for a cross-disciplinary Research Experience for Undergraduates site sponsored by NSF. He is currently the PI for an NSF project on Innovation in Graduate Education. Dr. Ram is an evaluator for the Engineering Accreditation Commission of ABET.

Abstract

The Graduate Research Identity Development program (GRID) is an initiative in the College of Engineering at North Carolina A&T State University, sponsored by the National Science Foundation since 2019. The program offers seminar-type lectures supplemented with activities designed to help graduate students develop critical skills for research-based careers. The program is focused on graduate engineering students but is open to graduate students from all programs. Students also choose mentors from within and outside the university with the goal of increasing their sense of belonging to the field and their identities as research engineers. As part of this program, a pilot study is in progress, aimed at performing a full-scale network analysis of student interactions. A web-based survey was administered to collect information about students in and outside the College of Engineering who participate in the GRID program sessions. The survey was designed to collect information on the relationship networks (or lack thereof) that students are involved in as they matriculate through their graduate program. It assesses things such as how and where the students interact with one another, members of faculty and staff, and with contacts from intramural and extramural organizations. Several items are also used to assess students' perceptions of themselves as *research engineers*. In this paper, we focus on the interactions of students in the classroom. More specifically, we form networks based on the student answers about the classes they have taken in different departments. We then analyze the resultant networks and contrast certain graph theoretic properties to students' scores on the research engineer identity items. Do students that are in the periphery, or students that have more connections attain higher research engineer identity scores? Do students that form complete subnetworks (cliques) or core-periphery structures (induced stars) have higher scores than others? This paper presents the findings from this pilot study from the network analysis on this cohort of students. In summary, we find that students with high eigenvector centrality scores and those who form larger cliques possess significantly higher research engineer identity scores.

Introduction

Since 2019, North Carolina A&T State University has been offering the Graduate Research Engineer Identity Development Program (GRID): a seminar-type series of lectures and activities that are designed to guide graduate students as they are building their research engineer identity. As part of the program, students were provided a web-based survey that provided us with information about how they connect to one another within and outside the university. In this pilot study, we aim to analyze the networks of interaction that result from their responses to the survey questions. We specifically want to address the following questions:

1. In visualizing the networks of students based on their coursework, do we obtain useful information about how they perceive themselves?
2. Specifically, is there a relationship between the network relationships between students, their bridging and bonding social capital, and their sense of identity as a research engineer?
3. Do students who participate in larger cliques also showcase higher levels of research engineer identity scores?

The remainder of the paper is organized as follows. First, we provide an overview of some of the fundamentals on social capital, communities of practice, and identity. Then, we present our methods. Namely, we discuss the survey that was conducted, the network that was generated, and provide our analysis and computational results. We conclude the paper with a brief discussion of the results and provide avenues for future work, with a more representative, larger sample of students.

Social Capital

Social capital refers to resources and information that are embedded in social networks (Lin 2001). As a community characteristic, social capital can impact the flow of other capitals, such as human and cultural capital. Communities are often described as having two different types of social capital. Bonding social capital involves strong ties that build cohesion, provide social support, and encourage loyalty within a relatively homogenous community, whereas bridging social capital involves loose ties that serve as connections between heterogeneous communities (Narayaun, 1999; Granovetter, 1973 & 1985). Whereas bonding social capital connects individuals who presumably already have access to similar resources, bridging social capital helps individuals access resources and information beyond what was available as a result of bonding social capital (Putnam 2000).

Recent research has applied the concepts of bridging and bonding social capital to understand students' experience completing their academic programs at universities. By conducting interviews with 37 students studying at Danish and Australian universities, researchers examined whether bonding and bridging social capital impacted academic and professional identity. They found that (1) students developing bonding social capital with other students facilitated academic identity formation but not professional identity formation; (2) students developing bridging social capital with educators facilitated both academic and professional identity formation; and (3) students developing bonding social capital negatively impacted their participation in efforts to develop bridging social capital with educators (Jensen and Jetten 2015). This latter finding suggests that there is a relationship between bonding and bridging social capital. Indeed, scholars have found that optimal outcomes occur when both bonding and bridging social capital are present (Saegert, Thompson, and Warren 2001; Warren et al. 2001; Stone and Hughes 2002).

Studies have had mixed findings regarding whether co-student interaction can develop bridging social capital. Whereas some researchers found that students did not develop bridging social capital from interactions with other students (Jensen and Jetten 2015), other researchers reported that minority student groups can obtain bridging social capital by engaging with majority or higher socioeconomic status students groups (Thomas 2002; Scanlon et al. 2007). Thus, more research is needed to explore the extent to which networks among students can develop bridging social capital.

Whereas some scholars have operationalized bonding social capital as co-student interaction and bridging social capital as student-faculty interaction (Jensen and Jetten 2015), we highlight the importance of examining bonding social capital among students and faculty within academic departments and bridging social capital among students and faculty across academic departments. Academic departments and learning cohorts within them can represent post-place communities that are "networks of people tied together by solidarity, a shared identity and set of norms, that does not necessarily reside in a place" (Bradshaw 2008). Feelings of connectedness within learning communities are important for reducing dropout rates among graduate students (Kraska 2008).

Access to social capital can be conceptualized as both a method of social control and a key to social mobility (Bourdieu 2001). Regardless of the intent, the impact that social capital has on key outcomes remains. Social capital is a predictor of academic persistence and educational attainment (Coleman 1988). The most disadvantaged populations tend to be the ones most spatially and socially isolated and with the most limited access to elites (Sherman 2006, Duncan 1999, Schulman and Anderson 1999, Lin 1999, Wilson 1978). This is important because elites are influential in connecting individuals to opportunities and providing access to top jobs and pay (Royster 2003).

In this paper, we explore the impact of bonding and bridging social capital on the development of research identity among graduate students pursuing research-based degrees. We define research identity as a personal, role-, and group-based "identity" as a researcher. Since social capital is a collective good, we hypothesize that:

- (H1) Students in departments with higher bonding social capital (e.g., form larger cliques, higher number of interactions) will have higher research identity.
- (H2) Students accessing bridging social capital will have higher research identity compared to students in the same department who are not accessing bridging social capital.
- (H3) Accessing bonding and bridging social capital in combination will produce higher research identity than either bonding social capital or bridging social capital alone. Students accessing bridging social capital in departments with lower bonding social capital will have lower research identity compared to students accessing bridging social capital in departments with higher bonding social capital.

Identity, Communities of Practice, and Social Capital

Efforts to improve STEM education in the US began in earnest in the early 2000s. Since this time, a burgeoning literature has developed on what types of structural, cultural, pedagogical and social psychological forces undergird STEM educational practices at both the K-12 and post-secondary levels. Among these efforts, many have worked to improve our understanding of how social psychological facets of being a STEM student may positively or negatively impact student learning outcomes and engagement. The construct of identity has quickly become one of the most useful tools for researchers and educators working in this area.

Identity is a rich construct with a long history of research and theory in Sociology, Psychology and many other fields. Research on identity within engineering education began in the early 2000s (Morelock, 2017). Here we draw primarily on Identity Theory and Social Identity Theory to define identity as meanings attached to the self-concept that position the self within networks of social relationships (Burke & Stets, 2010). This positioning includes seeing oneself as similar to some and different from others; (Walton & Jones, 2018). Individuals hold multiple identities all of which are dynamic, yet provide individuals a sense of consistency and stability by connecting the past with one's (perceived) future trajectory through providing an answer to the question, who am I? Importantly, the self-meanings that constitute one's identity are built up from social interactions and the reflected appraisals of others. It follows then that academic departments with stronger bonding capital (i.e., resources and information), likely provide students with more interactive opportunities to clearly define their identities as research engineers (H1). In similar fashion, by accessing the resources and information available via cross and interdisciplinary opportunities (i.e., bridging social capital), students increase the number of interactive opportunities available for them to clearly define their identities as research engineers (H2).

Wenger (1998) and others offer a theory of social learning grounded in the idea that learning or knowledge creation is grounded in regular social interaction within a domain of interest. Through social interaction and engagement in a shared practice, communities and knowledge are co-produced through the reification of ideas. In this way, learning (or knowledge construction) within Communities of Practice (CoP) occurs simultaneously along with identity and group development. Knowledge about a given domain (and one's understanding of that knowledge) emerge along-side the development of individuals, and groups. These processes constitute a co-development of individual, knowledge, and community.

Key to these processes are individuals' negotiation of their identities within the multiple communities that they engage with (Wenger, 2010). As noted above, identities provide individuals with a relatively stable

and consistent sense of "who they are" across time and social context. However, as we move between different communities of practice and are confronted with new and different experiences and knowledge systems, our definitions of self must adjust and be reconciled accordingly. This "...reconciliation may be the most significant challenge faced by learners who move from one community of practice to another" (Wenger, 2010, p.138). These experiences with the boundaries that distinguish one learning community from another can be rich sources of learning opportunities yet they may also present significant barriers to learning. Important to our study is that "the learning and innovation potential of a social learning system lies in its configuration of strong core practices and active boundary processes" (Wenger, 2010, p. 127). We argue that strong core practices are in part reflected in strong bonding social capital and that active boundary processes are in part reflective of strong bridging social capital. It follows then that students from departments with strong bonding capital (or strong core practices) and strong bridging capital (or active boundary processes), should have higher research engineer identities than students coming from departments lacking in either (H3).

Methods

In this section, we discuss the survey administered and its context, the measure of Research Engineer Identity that was devised and used, and the network analysis concepts employed. We finish the section with a discussion of the experimental setup: how we created the networks and how we obtained our findings.

Intervention

This paper stems from a National Science Foundation project to broaden the participation of underrepresented groups within graduate engineering programs at North Carolina A&T State University (NCAT), a leading HBCU and producer of African American engineers. The innovations in this project are three-fold: (1) a student-focused Research Engineer Network, (2) a faculty-focused formal Small Research Groups initiative, and (3) a Research Engineer Identity assessment process. The **Research Engineer Network (REN)** is a college-level initiative to develop skills and relationships to help graduate students seeking research-based degrees navigate the development of their research engineer identity. The cornerstone of the REN is a Graduate Research Engineer Identity Development Program that includes an eight-week collaborative workshop series covering five topics: (1) Research Progression Skills I (targeted for student's in the first year of their graduate program), (2) Research Progression Skills II (targeted for student's beyond the first year of their graduate program), (3) Research Networking Skills, (4) Career Preparation and Previews (led by research-leaning corporations and federal labs), and (5) mentoring for research identity. The **Small Research Groups** are groups of three to five faculty who have related research interests. This initiative seeks to address the vulnerabilities of single faculty research groups, that are common in small engineering colleges such as NCAT. The **Research Engineer Identity (REI) assessment process** involves the development and testing of an approach to assess research engineer identities among graduate students. The assessments include both formative and summative evaluations, the culmination of which will be a quasi-experimental paired-testing analysis in which the research engineer identities of a control group (students not participating in the REN), will be compared to an experimental group (students participating in the REN). This information will inform the development of a survey-based Research Engineer Identity Scale (REIS).

Measure Development

As part of the REI assessment process, a mixed-methods research design was used to develop a measure of Research Engineer Identity. First, a series of seven semi-structured focus groups were conducted with a diverse group of research engineers. Each focus group included about 6 to 9 graduate students, faculty

members, or industry professionals who actively engaged in engineering research in the Southeastern United States. The focus groups were designed to identify the content, character, and complications associated with efforts to develop a Research Engineer Identity. A total of 51 research engineers participated in the focus groups. The focus groups were audio recorded and transcribed. Next, two researchers used open coding (Esterberg, 2002) to identify key emerging themes within the focus group discussions. Following best practice, we then developed a codebook, had two coders use the codebook to conduct focused coding of all focus group transcripts, assessed interrater reliability across coders, and addressed all discrepancies in coding across coders. Content analysis was then conducted to analyze and identify patterns within and across codes (Mayring, 2010). Each of the underlying themes related to the self-meanings associated with being a research engineer (i.e., Research Engineer Identity), were then transformed into a closed-ended survey-based measure. A pool of 36 items were developed to assess Research Engineer Identity, and a subset of six of these items were used as a measure of Research Engineer Identity in the current pilot study[1]. Principal Components Analysis was used to produce component scores for these six items to serve as the dependent variable in the network analysis.

Survey Administration and Response Rate

We pilot tested the Research Engineer Identity Scale (REIS) using an online survey. In addition to the REIS, the online survey asked graduate students pursuing research-based degrees about their: (1) research identity, values, and efficacy, (2) behavioral intentions for next academic semester, (3) awareness, perceived need and participation in the program, (4), educational barriers they face, and (5) their research network orientation. We disseminated the online survey to a convenience sample of graduate students who were participating in research-intensive degree programs. We advertised the survey in three waves. In Wave 1, we disseminated the survey to graduate students who participated in a student-initiated mentoring program that connects students to faculty members and industry experts outside of their required thesis or dissertation committee (N=19 respondents; 95% response rate). In Wave 2, we disseminated the survey to graduate students who participated in the GRID program in the Fall of 2020 (N=26 respondents; 79% response rate). In Wave 3, we emailed a sample of graduate students who matched the demographic characteristics of respondents in Wave 1 & 2, but who did not participate in the GRID program (N=41 respondents; 35% response rate). The total responses across the waves were 86 graduate students, for an overall response rate of 51% across the three waves. We received institutional review board approval before survey administration.

Analytic Technique to Examine Centrality and Structures

Network analysis has been an area of research with wide-reaching implications for disciplines ranging from biology to the social sciences (Borgatti et al. 2009), as well as in traditional engineering fields (e.g., transportation networks, telecommunications networks). For defining a network, we simply need to define two sets: a set of nodes (entities) and a set of edges (revealing whether the nodes/entities interact). When two nodes do not immediately interact, then they may connect to one another through a path using other, intermediary nodes. The path that connects two nodes using the smallest number of intermediaries (possibly zero) is referred to as a shortest path.

Naturally, as soon as we obtain a network representation of the phenomenon we are studying, we may want to classify entities as more important or less important. Such studies of importance typically focus

---

[1] The psychometric properties of the full Research Engineer Identity (REI) scale are currently being evaluated so data are not reported here.

on notions of *centrality* (Wasserman and Faust 1994): i.e., how central is a certain entity in the network compared to another. Typical measures of centrality include:

1. degree centrality, which becomes higher as the number of connections that a node has in the network increases.
2. betweenness centrality, which is higher when the number of shortest paths that use a node as an intermediary over all available shortest paths increases.
3. eigenvector centrality, which extends degree centrality by including that connections towards other important nodes count more.

Based on our earlier discussion, degree centrality can be viewed as being analogous to bonding social capital and core practices; betweenness as bridging social capital and boundary processes. Finally, eigenvector centrality can be viewed as a combination of the two, favoring nodes that would showcase both bonding (core practices) and bridging (boundary processes) social capital.

In a network, we may also want to search for specific types of structures. Two extremes are cliques (complete subgraphs, which can be used as a proxy to bonding social capital and active core practices) and induced stars (a center node adjacent to numerous leaf nodes, which can be used as a proxy to bridging social capital and active boundary processes). With the term clique, we mean a set of nodes that share all possible pairwise connections; with the term (induced) star, we imply a set of nodes each with a connection to a common center and no other edges between other pairs of nodes. To find the largest induced clique that includes a certain node and the largest induced star that includes that node as a center, we solve the integer programs presented in (1) and (2), respectively. We also present pictorial examples of cliques and stars in Figure 1 in (a), (b), (c) and (d), (e), (f), respectively.

In both mathematical formulations, x_i is a binary variable that is equal to 1 if and only if node i is in the clique/star. Additionally, in the clique formulation, two nodes are not allowed to both be in the clique, unless they share an edge. On the other hand, in the star formulation, two non-center nodes are now allowed to both be in the star if they share an edge.

$$
\begin{aligned}
\max \quad & \sum_{i \in V} x_i \\
\text{s.t.} \quad & x_i + x_j \leq 1, \quad \forall (i,j) \notin E, \\
& x_u = 1, \\
& x_i \in \{0,1\}, \quad \forall i \in V.
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
\max \quad & \sum_{i \in V} x_i \\
\text{s.t.} \quad & x_i + x_j \leq 1, \quad \forall (i,j) \in E : i \neq u, j \neq u, \\
& x_i = 0, \quad \forall i \in V : (i,u) \notin E, \\
& x_u = 1, \\
& x_i \in \{0,1\}, \quad \forall i \in V.
\end{aligned} \tag{2}
$$

(a) A clique of size 5. (b) A clique of size 10. (c) A clique of size 3.

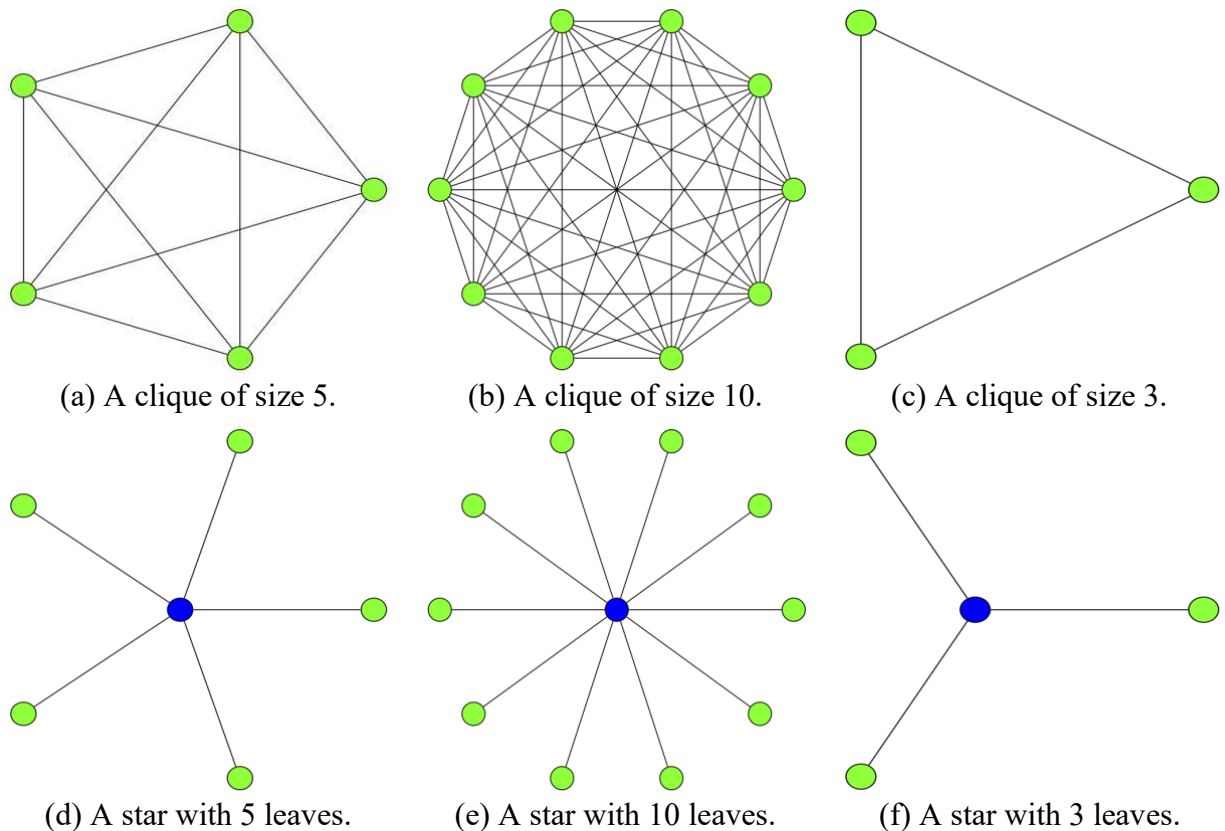(d) A star with 5 leaves. (e) A star with 10 leaves. (f) A star with 3 leaves.

Figure 1. Six examples of cliques and stars. The cliques presented in (a), (b), and (c) represent 5, 10, and 3 nodes that are all adjacent to each other. A clique of size 3 is also referred to as a triangle. The stars presented in (d), (e), and (f) represent graphs where all leaves are adjacent to a center (depicted in blue). The number of leaves are 5, 10, and 3.

Network Generation

For generating our network $G(V, E)$, we focus on a specific part of the questions in the survey: the courses graduate students have taken from each department during their tenure at the university. The set of nodes of the network, V, includes all students who responded to the course questions. Then, two students are connected by an edge in E if they have taken at least l classes from one department in common. As an example, consider a student that has registered for 2 classes and another student that has registered for 3 from Industrial and Systems Engineering. Then, these two students would be connected by an edge for $l \leq 2$, but would not share an edge for $l > 2$.

In the remainder of our experiment, we will investigate different resultant networks for $l \in \{1, 2, 3, 4, 5, 6, 7\}$ with the corresponding networks presented in Figure 2. We do not investigate higher numbers than this, as they lead to a single set of students (all from the same department) who have taken at least 8 classes from the same department (see Figure 3). Due to the limited number of returned surveys with coursework information, as well as the fact that this is a pilot study, we ended up with several participants who had no interactions with the rest of the survey respondents. This is why we present solely the *largest connected component* of each graph: that is, the largest number of nodes that are connected by a path.

We continue with a visualization of the networks obtained. All network analyses and visualizations were performed using Python and networkx (Hagberg et al. 2008). We show all of the course networks for different values of l in Figure 2. The numbers that appear as a label to each node/student represent the values of the research engineering identity metric value.

We immediately observe two things. The networks for $l = 4, 5$ are exactly the same; the same is true for the networks obtained for $l = 6$ and $l = 7$. Then, we report three graph theoretic measures based on node centrality and two measures based on the biggest structure in which the students participate. Namely, we first report the degree, betweenness, and eigenvector centrality for each of the nodes. Secondly, we report the largest induced clique that each node participates in, and the largest induced star that the node serves as a center for. Finally, we use these values to investigate whether a linear relationship exists between the centrality metrics and the size of clique/star against the research engineering identity scores, as measured by the answers in the survey. More specifically, we perform a simple linear regression with the centrality metrics and the size of the largest clique/star of a specific node as the independent variable and the research engineer identity score as the dependent variable. We then report the AIC scores for each of the regression models and for all networks constructed with lower AIC scores reflecting a better and more parsimonious fit to the data.

As an example, consider the $l = 3$ network presented in Figure 2 (c). Each of the connections present in the network signals a pair of students that have shared at least 3 courses from the same department in the College of Engineering. This results in an easier distinction between departments (e.g., in the lower left corner, we can identify a group of Industrial and Systems Engineering students). It also results in a set of students that are interdisciplinary, having taken at least 3 courses from other departments, too.

Computational Results

For the six networks generated (one for each l parameter value), we obtain the nodal centrality metrics and clique/star sizes (as described earlier) and search for their linear relationship to the research engineering identity score. Our results are summarized in Tables 1 and 2. The results are also presented as box plots in Figures 4—6 (for the three centrality metrics) and Figures 7—8 (for the clique and star sizes).

From the results in Table 1 and in Figures 4, 5, and 6, we may deduce two things. As l increases in number, all three centrality metrics improve in their capacity to discriminate between lower and higher research engineer identity metrics. That said, betweenness centrality (perhaps surprisingly) never becomes as good a discriminant as degree (to an extent) and eigenvector centrality. The second thing we observe is that eigenvector centrality outperforms the other centrality metrics quite significantly, as revealed by our AIC scores. It consistently leads to a better fit between the value it attains and the identity metric value: in summary, eigenvector centrality better and more parsimoniously models the identity metric value obtained.

As mentioned in the previous paragraph, perhaps surprisingly, students with high betweenness centrality (i.e., students that serve as bridges between departments, taking multiple classes from multiple departments, students that serve as connectors for different departments) showcase no significantly higher research engineering identity metric. This is an interesting result that we look forward to investigating more with a larger, more representative sample of survey responses.
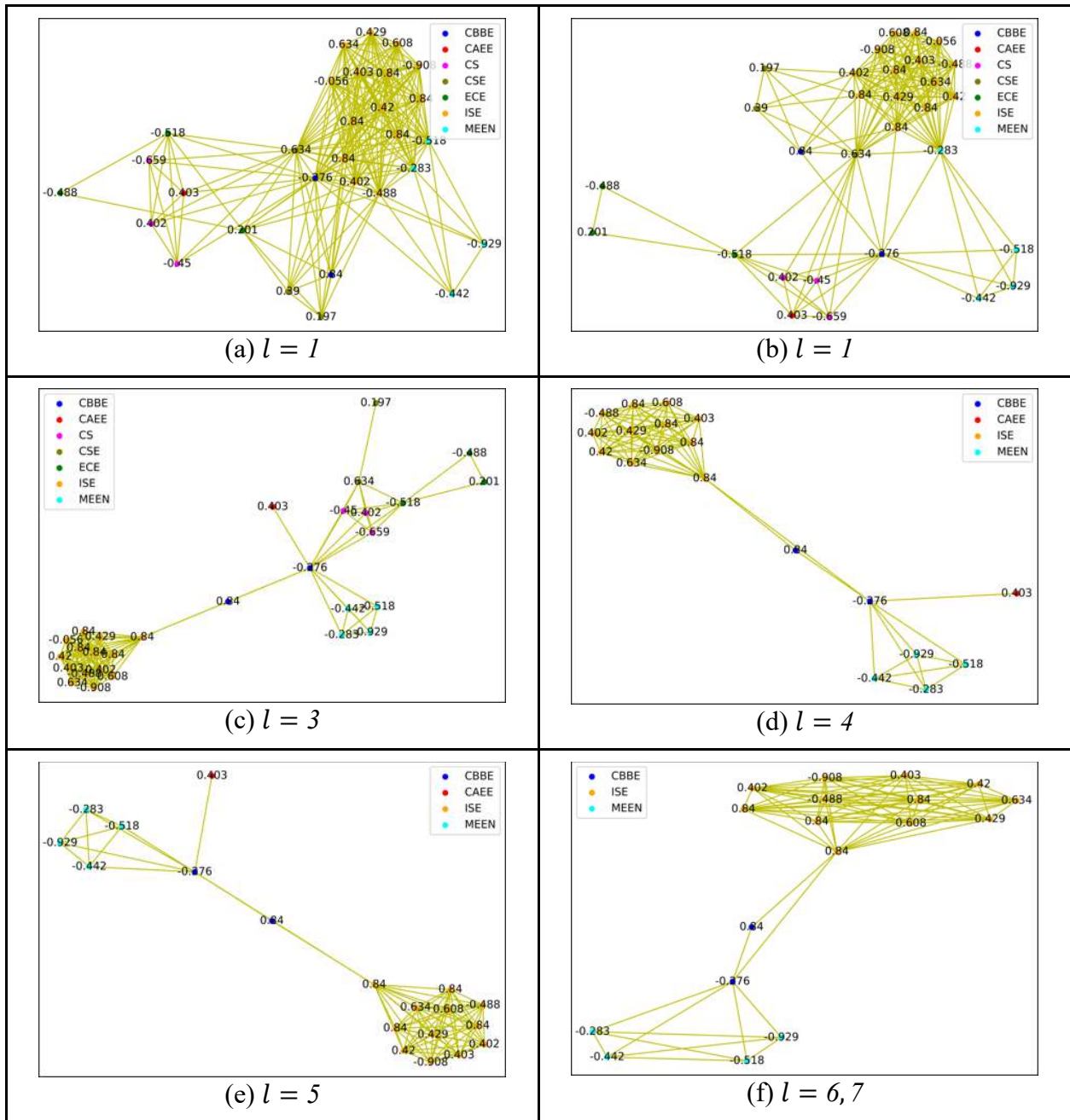
Figure 2. The three class networks for $l = 1, ..., 7$. We observe how the network is expectedly getting sparser as the requirement of interaction becomes harder to satisfy. We also observe that the groups of students interacting within their "own" departments (i.e., for the majors they pursue) get clearer and easier to separate for $l \geq 3$.
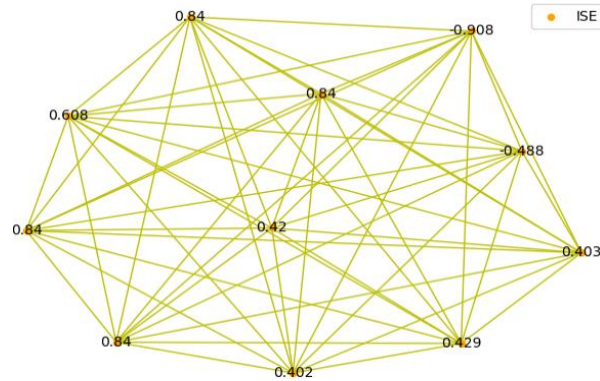
Figure 3. The network obtained for $l = 8$. It is a complete subgraph (clique) consisting of 11 Industrial and Systems Engineering students. Because of that, it was excluded from our analysis.

Network	Betweenness centrality AIC	Degree centrality AIC	Eigenvector centrality AIC
$l = 1$	56.36	52.75	51.23
$l = 2$	55.44	48.99	47.37
$l = 3$	54.52	50.53	47.50
$l = 4, 5$	38.83	36.22	34.08
$l = 6, 7$	38.11	33.61	31.97

Table 1. A summary of the AIC scores obtained when checking for a relationship between the three centrality metrics and the research engineering identity construct value for the networks generated.
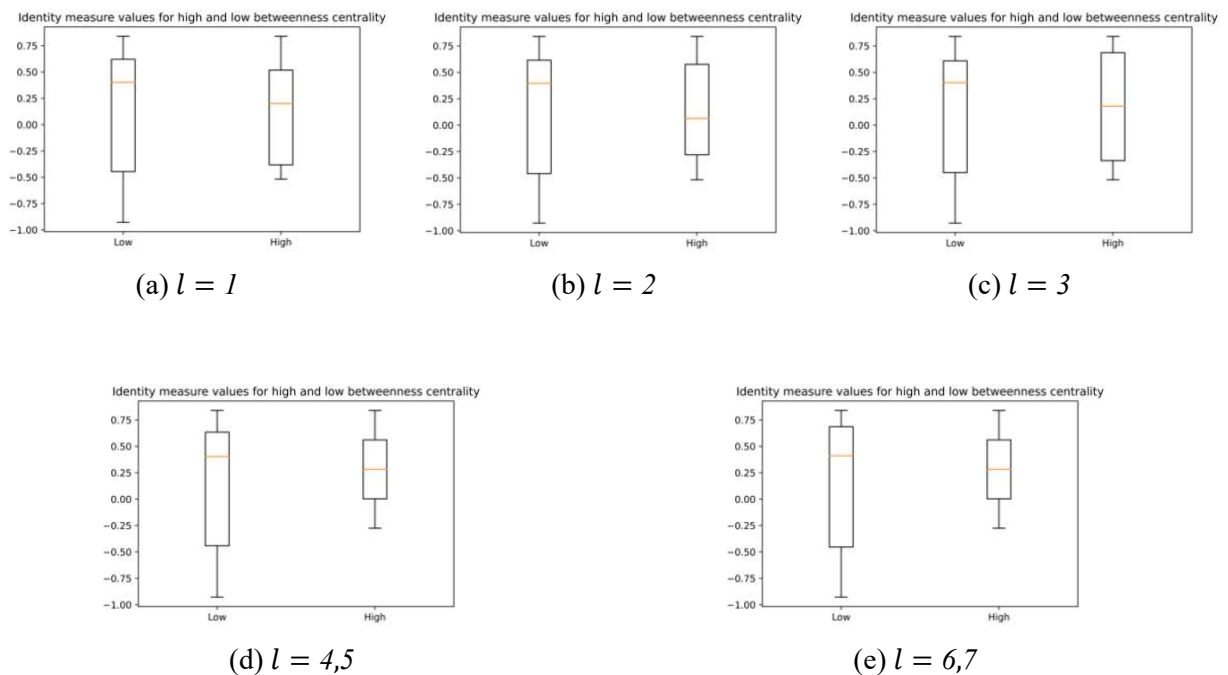

(a) $l = 1$ (b) $l = 2$ (c) $l = 3$

(d) $l = 4,5$ (e) $l = 6,7$

Figure 4. A set of box plots to show the research engineering identity metric and how it varies between nodes with high and low betweenness centrality metric values for $l = 1, ..., 7$. For our purposes here a node is considered to have high betweenness centrality if it is above the mean betweenness centrality value.

(a) $l = 1$ (b) $l = 2$ (c) $l = 3$

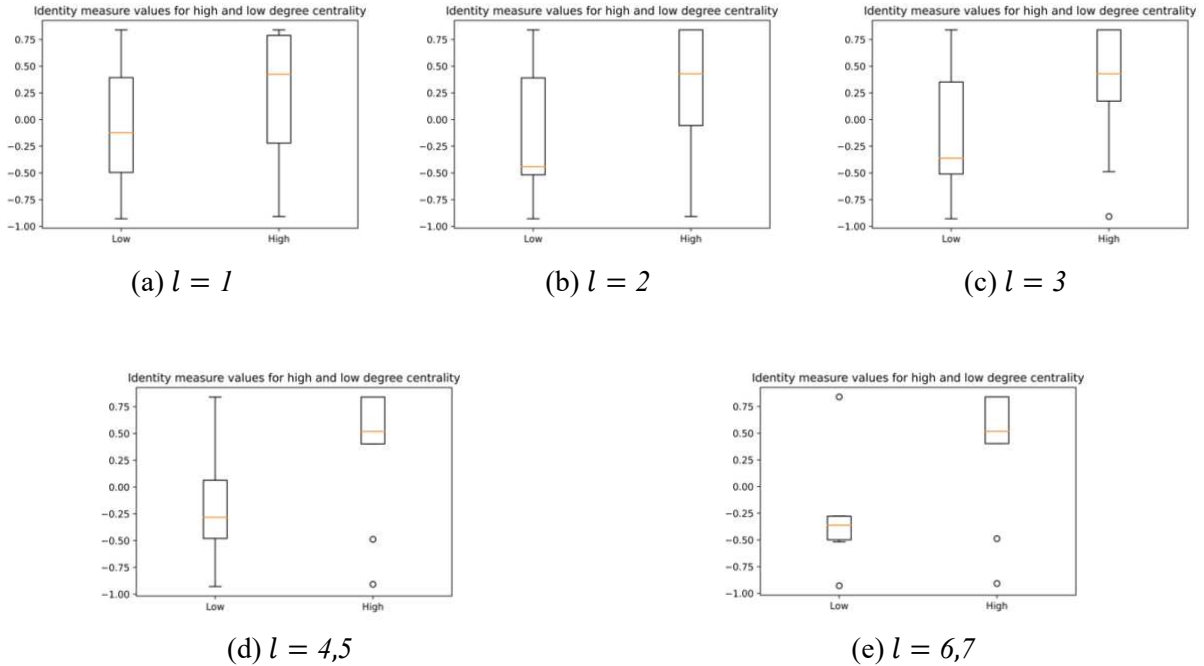(d) $l = 4,5$ (e) $l = 6,7$

Figure 5. A set of box plots to show the research engineering identity metric and how it varies between nodes with high and low degree centrality metric values for $l = 1, ..., 7$. For our purposes here a node is considered to have high degree centrality if it is above the mean degree centrality value.

(a) $l = 1$ (b) $l = 2$ (c) $l = 3$

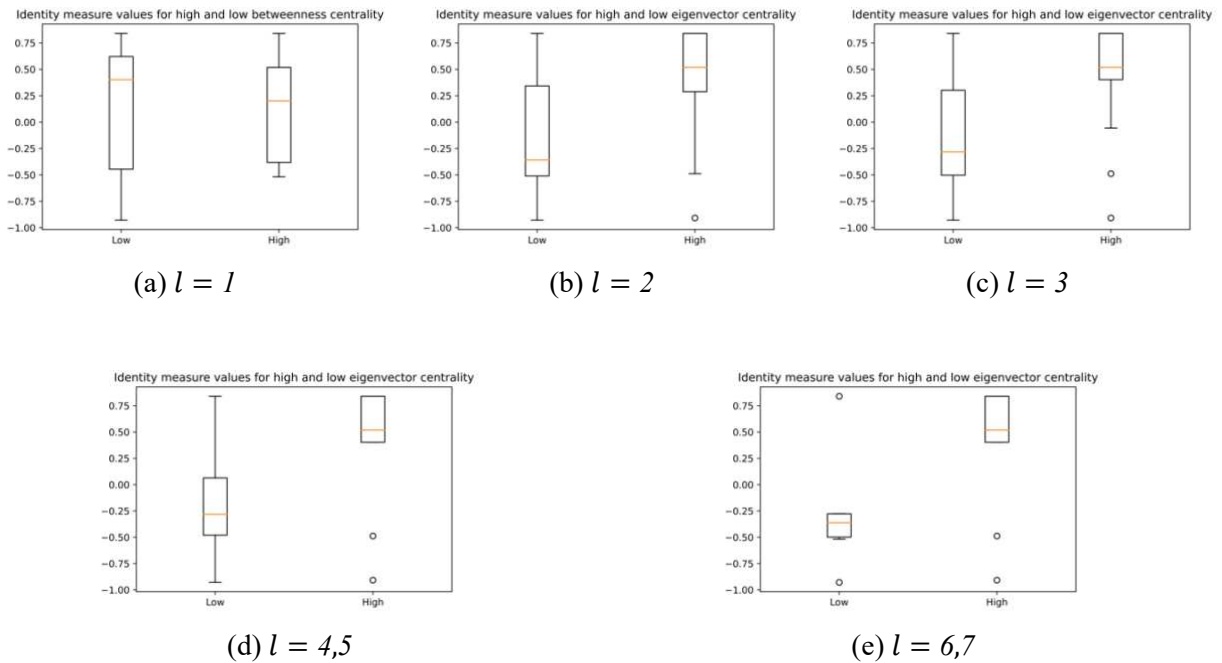(d) $l = 4,5$ (e) $l = 6,7$

Figure 6. A set of box plots to show the research engineering identity metric and how it varies between nodes with high and low eigenvector centrality metric values for $l = 1, ..., 7$. For our purposes here a node is considered to have high eigenvector centrality if it is above the mean eigenvector centrality value.

Another item we wanted to investigate is whether the largest structure that a student participates in affects their research engineering identity metric score. To do that, we identified the largest clique structure that contains the students, and the largest star structure that includes the student as its center. Our results, summarized in Table 2 and presented as box plots in Figures 7 and 8, seem to indicate that these two do not play a significant factor in whether students score higher in the research engineering identity metric.

Network	Star size AIC	Clique size AIC
$l = 1$	55.35	50.51
$l = 2$	54.85	48.68
$l = 3$	54.42	49.77
$l = 4, 5$	38.87	36.20
$l = 6, 7$	37.87	33.86

Table 2. A summary of the AIC scores obtained when checking for a relationship between the size of the largest clique containing a node or the size of the largest star centered at a node and the research engineering identity construct value.
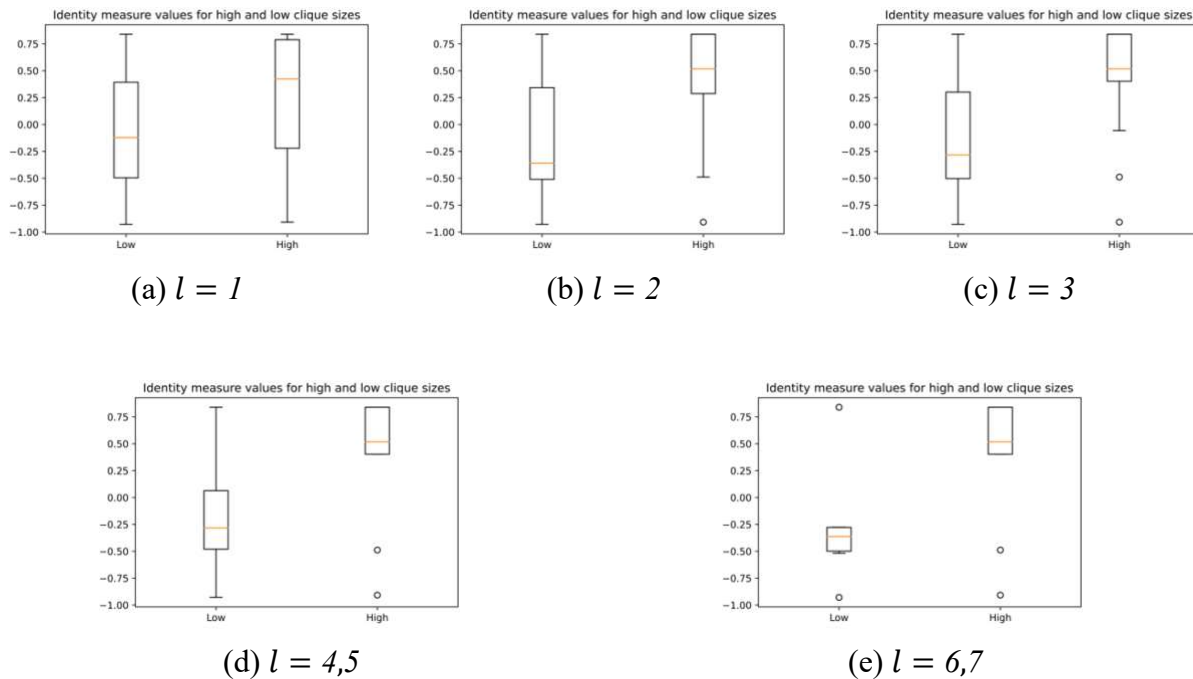

(a) $l = 1$ (b) $l = 2$ (c) $l = 3$

(d) $l = 4,5$ (e) $l = 6,7$

Figure 7. A set of box plots to show the research engineering identity metric and how it varies between nodes that participate in high and low size cliques. For our purposes here a node is considered to have high clique size if it is above the mean clique size value.

Based on the AIC scores obtained, we observe that clique size leads to a better and more parsimonious fit for modeling research engineering identity than star size. Furthermore, information presented in Figure 8 indicates that star size may be an unimportant factor in determining research engineer identity.
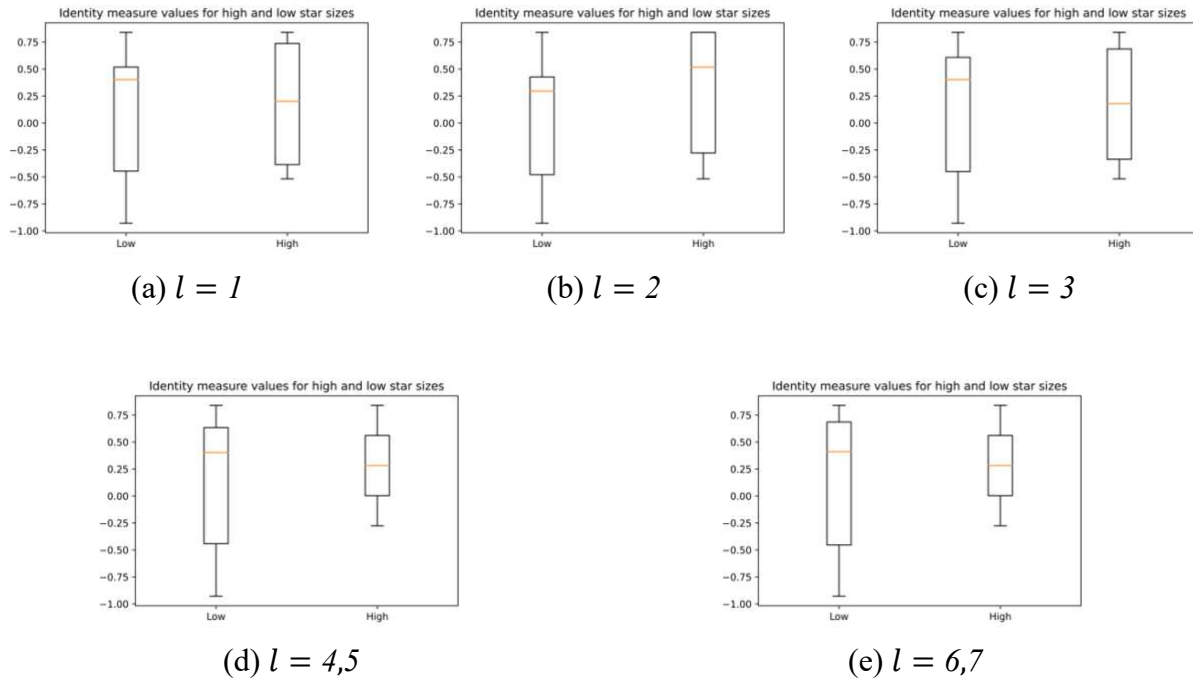
(a) *l = 1*　　　　(b) *l = 2*　　　　(c) *l = 3*

(d) *l = 4,5*　　　　(e) *l = 6,7*

Figure 8. A set of box plots to show the research engineering identity metric and how it varies between nodes that participate in high and low size stars. For our purposes here, a node is considered to have high star size if it is above the mean star size value.

Discussion and Future Research

The most important finding from this pilot study is that we are able to discern which students are more likely to showcase higher research engineer identity metric values from their coursework network. More specifically, we observed that students with high eigenvector centrality metrics also showcased higher research engineer identity metric values. This became clearer for larger values of the threshold of the shared coursework number from a department. This makes sense, as for smaller values of this threshold, even students who had only shared fewer classes would appear connected in the network, confounding our results. This finding provides support for our third hypothesis that accessing bonding and bridging social capital in combination will produce higher research identity than either bonding social capital or bridging social capital alone. This finding is consistent with previous studies that found optimal outcomes occur when both bonding and bridging social capital are present (Saegert, Thompson, and Warren 2001; Warren et al. 2001; Stone and Hughes 2002). This finding also supports work on communities of practice. As noted earlier, strong core practices within a community of practice and active boundary processes between communities of practice are thought to and have been shown to promote learning and innovation potential (Wenger, 1998; Ingram et al., 2014)

Additionally, we observed that degree centrality also served as a better model of research engineer identity than betweenness centrality. This was supported by our reported AIC values comparing the relative importance of degree and betweenness centrality. Relatedly, we noticed that students who form larger cliques when considering the number of courses they have taken from each department are also more likely to showcase higher research engineer identity scores. These findings provide support for our first hypothesis that students in departments with higher bonding social capital will have higher research identity. Findings are inconsistent with Jensen and Jett's (2015) study that found students having bonding social capital facilitated academic identity formation but not professional identity formation. Instead, findings lend support to Thomas (2002) and Scanlon et al. (2007), which highlight co-student interaction

as important for developing bonding social capital. Central to the learning community literature is the idea that feelings of connectedness are important to a range of positive outcomes, including academic persistence and identity development. Consistent with theories of learning communities, we find that bonding social capital is both a necessary and sufficient condition to produce research identity.

On the other hand, betweenness centrality which is akin to bridging social capital, was a less important factor and we were unable to discern between high and low research engineer identity scores using betweenness. Related to this finding, induced stars were not related to research engineer identity scores. Previous studies found that bridging social capital was important for professional identity formation (Jensen and Jetten 2015). In contrast, our study provides nuance by illustrating that bridging social capital on its own is not necessary or sufficient to achieve research identity. Yet, bridging social capital can enhance research identity development when bonding social capital is present. Each of these findings align well with work on communities of practice. Indeed, experiences in which one interacts with other learning communities outside of one's own (e.g., bridging), force individuals to reconcile their identities and can promote enhanced learning and identity development, but also present powerful barriers to these outcomes (Wenger, 1998). Our research seems to suggest that strong core practices within a community that provide bonding social capital may provide a needed foundation for cross community learning and identity development.

Seeing as this is a pilot study, there are a lot of avenues to explore in the near future. The first question that we need to address is whether these findings extend to different types of networks that can be generated: will we be seeing this relationship between centrality and structures, and the research engineer identity scores in student friendship networks, in student-faculty relationship networks, or in intramural and extramural organization participation networks? Additionally, what do we observe when we consider the centrality of classes (Everett and Borgatti 1999) of students, such as students in the same department, or students with the same advisors? In a similar vein, we would like to investigate how our results generalize when we look into the centrality of groups that are supposed to induce specific structures, such as cliques, stars, representatives, among others (Rasti and Vogiatzis 2021).

References

Agnitsch, K., Flora, J., and Ryan, V. (2006). Bonding and bridging social capital: The interaction effects on community action. *Community Development 37*:36-51.

Bourdieu, P. (2001). The forms of capital. Pp. 96-111 in *The Sociology of Economic Life*, edited by M. Granovetter and R. Swedberg. Boulder: Westview Press.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *science*, *323*(5916), 892-895.

Bradshaw, T.K. (2008). The post-place community: Contributions to the debate about the definition of community. *Journal of the Community Development Society 39*: 5-16.

Burke, P. J., & Stets, J. E. (2009). Identity Theory. New York, NY: Oxford University Press.

Coleman, J. S. (1988). Social capital and the creation of human capital. *American Journal of Sociology 94:*S95-120.

Duncan, C. M. (1999). *Worlds Apart: Why Poverty Persists in Rural America.* New Haven: Yale University Press.

Esterberg, K. G. (2002). *Qualitative methods in social research.* Boston, MA: McGraw-Hill.

Everett, M. G., & Borgatti, S. P. (1999). *The centrality of groups and classes.* The Journal of mathematical sociology, 23(3), 181--201.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology 78*: 1360-1380.

Granovetter, M. S. (1985). Economic action, social structure and embeddedness. *American Journal of Sociology 91*:481-510.

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using NetworkX*(No. LA-UR-08-05495; LA-UR-08-5495). Los Alamos National Lab.(LANL), Los Alamos, NM (United States).

Ingram, J., Maye, D., Kirwan, J., Curry, N., & Kubinakova, K. (2014). Learning in the permaculture community of practice in England: an analysis of the relationship between core practices and boundary processes. The Journal of Agricultural Education and Extension, 20(3), 275-290.

Jensen, D.H., and Jetten, J. (2015). Bridging and bonding interactions in higher education: social capital and students' academic and professional identity formation. *Frontiers in Psychology.* https://doi.org/10.3389/fpsyg.2015.00126

Kraska, M. (Fall 2008). Retention of graduate students through learning communities. Retrieved March 4, 2021, from http://scholar.lib.vt.edu/ejournals/JITE/v452/kraska.html.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.

Lin, N. (1999). *Building a network theory of social capital*. Connections 22:28-51.

Mayring, P. (2010). *Qualitative content analysis.* In U. Flick, E. von Kardoff, & I. Steinke (Eds.), *A Companion to Qualitative Research.* Thousand Oaks, CA: Sage.

Morelock, J. R., (2017) A systematic literature review of engineering identity: definitions, factors, and interventions affecting development, and means of measurement. *European Journal of Engineering Education,* 42(6), 1240-1262

Narayan, D. (1999). *Bonds and Bridges: Social Capital and Poverty.* Washington: World Bank, Report no. 2167.

Bomze, I. M., Budinich, M., Pardalos, P. M., & Pelillo, M. (1999). The maximum clique problem. In *Handbook of combinatorial optimization* (pp. 1-74). Springer, Boston, MA.

Putnam, R. D. (2000). *Bowling Alone: The Collapse and Revival of American Community.* New York: Simon-Schuster.

Rasti, S., Vogiatzis, C. (2021). Novel centrality metrics for studying essentiality in protein-protein interaction networks based on group structures, Networks, to appear.

Royster, D. S. (2003). *Race and the Invisible Hand: How White Networks Exclude Black Men from Blue Collar Jobs.* Berkeley, CA: University of California Press.

Saegert, S., Thompson, J.P., and Warren, M. R. (2001). *Social Capital and Poor Communities.* New York: Russel Sage Foundation.

Scanlon, L., Rowling, L., and Weber, Z. (2007). 'You don't have like an identity... you are just lost in a crowd': forming a student identity in the first-year transition to university. *J. Youth Stud 10*:223–241.

Schulman, M., and Anderson, C. (1999). The dark side of the force: A case study of restricting and social capital. *Rural Sociology 64:*351-372.

Sherman, J. (2006). Coping with rural poverty: Economic survival and moral capital in rural America. *Social Forces 85*:891-913.

Stone, W., and Hughes, J. (2002). Why social capital matters. *Family Matters*, Winter 62-67.

Thomas, L. (2002). Student retention in higher education: the role of institutional habitus. *J. Educ. Policy 17*:423–442.

Walton, T. N., & Jones, R. E. (2018). Ecological identity: The development and assessment of a measurement scale. *Environment and Behavior*, *50*(6), 657-689.

Warren, M.R., Thompson, J.P., and Saegert, S. (2001). The role of social capital in combating poverty. In S. Saegert, J.P. Thompson, and M. Warren (Eds.) *Social Capital and Poor Communities.* New York: Russel Sage Foundation.

Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications.

Wenger, E. (1998) Communities of practice: learning, meaning and identity. New York: Cambridge University Press, extracts reproduced with permission.