Formation, Measurement, and Imputation of Social Ties

by

Jordan T. Bates

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2019 by the
Graduate Supervisory Committee:

Spiro J. Maroulis, Co-Chair
Yun Kang, Co-Chair
Kenneth A. Frank

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

Network analysis is a key conceptual orientation and analytical tool in the social sciences that emphasizes the embeddedness of individual behavior within a larger web of social relations. The network approach is used to better understand the cause and consequence of social interactions which cannot be treated as independent. The relational nature of network data and models, however, amplify the methodological concerns associated with inaccurate or missing data. This dissertation addresses such concerns via three projects. As a motivating substantive example, Project 1 examines factors associated with the selection of interaction partners by students at a large urban high school implementing a reform which, like many organizational improvement initiatives, is associated with a theory of change that posits changes to the structuring of social interactions as a central causal pathway to improved outcomes. A distinctive aspect of the data used in Project 1 is that it was a complete egocentric network census – in addition to being asked about their own relationships, students were asked about the relationships between alters that they nominated in the self-report. This enables two unique examinations of methodological challenges in network survey data collection: Project 2 examines the factors related to how well survey respondents assess the strength of social connections between others, finding that "informant" competence corresponds positively with their social proximity to target dyad as well as their centrality in the network. Project 3 explores using such third-party reports to augment network imputation methods, and finds that incorporating third-party reports into model-based methods provides a significant boost in imputation accuracy. Together these findings provide important implications for collecting and extrapolating data in research contexts where a complete social network census is highly desirable but infeasible.

ACKNOWLEDGMENTS

Looking back on my journey to this point, I am overwhelmed with gratitude. I have been incredibly blessed to receive such depths of support and encouragement from so many – teachers, mentors, friends, and family.

This dissertation would not have been possible without the steadfast support of my advisor Spiro Maroulis. It has been a great pleasure and a privilege to work with him. I am grateful for his incredible selflessness, patience, and empathy. His positivity and encouragement kept me going when I felt discouraged, and our stimulating conversations never failed to brighten my day. I have been awed by his generosity and tireless advocacy. It is thanks to his wisdom and guidance that I made it to the finish line. I could not have asked for a better advisor.

I thank my co-chair Yun Kang and committee member Ken Frank for their time and helpful feedback. I am grateful to Carlos Castillo-Chávez and Erik Johnston for their unfettered support and for encouraging me to explore my interests not only with regards to the line of research leading to this dissertation, but also my various data science for social good pursuits. Thank you to Sherry Woodley and Margaret Murphy-Tillis who helped shepherd me through it all.

I am lucky to have had many wonderful mentors who gave me the opportunity to participate in research as an undergraduate, encouraged my pursuit of graduate studies, and supported me in so many other ways. I am particularly grateful for the mentorship of Jennifer Neville, Eugenio Culurciello, Susanne Hambrusch, Jeffrey Karpicke, Lynne Reder, and Stephanie Tristram-Nagle. I owe so much of my success to you. I also credit the excellent teachers who encouraged my curiosity, especially my math teachers Shadd Clarke, Brian Piper, Deborah White-Dunlap, and Patti Moss.

Thanks to all of my colleagues and friends for the camaraderie and good cheer these past six years. To all of the MCMSC crew, especially Dustin Padilla, Oyita Udiani, Victor Moreno Martínez, and Komi Messan. To my superheroes Pat Pataranutaporn, Kacie McCollum, Ram Polur, and Bin Hong Lee. To my long time friends Jannah Armstrong, Greg Thompson, Henry Char, and Max Malone. A mis amigas Andrea Fernández Conde y Ana Bertha Cruz Martínez. Les agredezco por ayudarme a disfrutar las cosas buenas que tiene la vida. My life has been made immeasurably richer for knowing all of you.

Finally, I thank my family. My mother Loree and my father Bill supported me from the beginning, picked me up when I fell, and gave me the gift of unconditional love. My sister Raven and brother-in-law Travis offered refuge and respite from my academic struggles. I will forever cherish the memories of our travels and time together these past few years. My brother Aiden gave me the motivation to push through this final stretch. I am delighted to be with him as I write these words. I dedicate this dissertation to Martha. I could not have wished for a more patient and supportive partner. Thank you, my love.

This dissertation was funded in part by the U.S. Department of Education through the Graduate Assistance in Areas of National Need (GAANN) grant program and by the Simon A. Levin Mathematical, Computational and Modeling Sciences Center.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

LIST OF TABLES

Table	Page

LIST OF FIGURES

EXTENDED ABSTRACT

Project 1 models the selection of interaction partners utilizing unique longitudinal social network data from a cohort of students in a small learning community situated within a larger urban high school. Reducing segregation on academic achievement was a key goal of school reform movements to racially integrate schools and to remove explicit tracking. Despite these efforts academic segregation has persisted in the form of implicit tracking. The small schools reform does away with implicit tracking, but students might still self-select into segregated social groups. No study to date has used social network analysis to examine friendship selection within this context. The results show that despite the fact that the students in the small learning community are integrated in the sense that they take the same classes together, their friend groups are remain segregated by academic achievement.

Project 2 explores the role of third-party evaluations in network measurement. Better understanding of the relationship between third-party and self-reports of social relations may help us to more accurately and efficiently measure social networks. I compare self-report to third-party reports of social ties among the students in the small learning community studied in Project 1. The students were surveyed with a complete egocentric network census – that is, all the students were first asked to list their own ties and then asked to report on the ties between those that they had nominated in the self-report. Using this data I develop and estimate a model of informant competence. I find that informant competence corresponds positively with academic achievement and that girls are better informants than boys. Furthermore, I find that informant competence corresponds positively with measures of the informant's social proximity to the dyad in question as well as with measures of the informant's network centrality.

These results suggest that informants could be screened not only for their likely global competence but could also be selected jointly to optimize coverage such that all dyads are evaluated by informants likely to be knowledgeable about their relationship.

Project 3 examines the issue of missing data. When conducting a network census, it is desirable to have as close to complete response as possible. Due to the relational nature of the data itself and approaches to modeling it, missing data is more problematic than for studies that focus on the individual. Unfortunately, avoiding non-response entirely may be unrealistic, particularly in school settings where a parental consent form is required. The most common methods for dealing with missing data in networks are deletion-based and studies have shown that model-based methods perform better. Use of third-party reports as a method for handling missing links is less common. In this project I augment a model-based method with third-party information and find that it provides a significant boost in imputation accuracy. Researchers may wish to consider using third-party reports as well as model-based imputation to deal with missing self-report data, especially when non-response is high.

Chapter 1

PROJECT 1: MODELS OF FRIENDSHIP SELECTION

Academic segregation can occur between schools, between formal or informal tracks within schools, and between friend groups within schools. Reducing segregation on academic achievement was a key goal of school reform movements to racially integrate schools and to remove explicit tracking. Despite these efforts academic segregation has persisted in the form of implicit tracking. The small schools reform does away with implicit tracking, but students might still self-select into segregated social groups. No study to date has used social network analysis to examine friendship selection within this context. This project utilizes unique longitudinal social network data from a cohort of students in a small learning community situated within a larger urban high school. Results show that despite the fact that the students in the SLC are integrated in the sense that they take the same classes together, their friend groups remain segregated on academic achievement.

1.1 Introduction

The grand disparities in academic outcomes in the United States are an affront to our democratic values. Moody (2001) notes that school integration was driven by "the recognition that separate could never be equal, in part because the social relations formed in school are an essential part of the educational process". However, he found that despite of the formal integration of schools the lived experience of students within schools remains substantively segregated. Segregation on academic achievement contributes to the perpetuation of the achievement gap. When students are segregated into low- and high-achieving groups, teachers tend to set reduced expectations and to provide poorer quality instruction to the low-achieving group. Friends are also important in shaping educational outcomes. Friends provide social capital and establish norms. When students are segregated on academic achievement these social dynamics serve to increase the gap between the poorly performing students and the high achieving students. Recognition of these facts has motivated advocates for equity to promote reforms that seek to reduce segregation on academic achievement.

This project looks at the case of a small-schools reform at a large urban high school using longitudinal social network surveys. The aim is to see whether the data is consistent with claims of the impact of small schools reform on the formation of peer groups.

This introduction first describes the role that friendship networks play in academic outcomes. Second, it discusses relevant findings from studies of adolescent friendship formation. Finally, it discusses school reforms related to academic segregation and their impact on friendship networks.

1.1.1 Friendship Networks and Academic Outcomes

1.1.1.1 Network Composition

The impact of influence and selection in producing homophily on academic outcomes has been estimated using stochastic actor-based models (SABMs) and both processes were found to be significant (Lomi et al. 2011; Flashman 2012). Random assignments to college peer groups have allowed researchers to test for peer effects (social contagion) and they have found evidence for peer influence on academic achievement (Sacerdote 2000; Zimmerman 2003; Carrell, Fullerton, and West 2008). Peer effects had the greatest impact on college retention of all factors considered in a social network study by Eckles and Stradley (2012). However, the issue is not completely settled – some researchers still argue that peer effects on academic achievement are not significant (Foster 2006; Wentzel, Barry, and Caldwell 2004).

1.1.1.2 Network Structure

1.1.1.2.1 Network closure

Social integration has been found to be a key factor in academic achievement. One way to operationalize social integration is by ego network density which may also be called connectedness or closure. It is theorized that dense connections help to maintain strong social norms. Maroulis (2008) conducted the social network surveys that are used in the present study. He found that while network structure (ego network density) and network composition (lagged peer GPA) did not have statistically significant effects on academic performance individually, their interaction did have a significant positive effect.

1.1.1.2.2 Brokerage

In contrast to the network closure theory of social capital, another hypothesized source of social capital is brokerage in horizon expanding networks wherein students occupy structural holes between different groups. Gasevic, Zouaq, and Janzen (2013) found that eccentricity and closeness centrality have a significant positive association with GPA in a Master's program. Thomas (2000) found that multiple measures of network structure related to persistence:

1. Bonacich centrality and in-degree were both positively associated with persistence
2. an out-degree in the middle of the distribution was associated with higher persistence while low and high out-degree were both associated with lower persistence
3. the percentage of ties falling within one's peer group was negatively associated with persistence.

1.1.2 School Reforms, Academic Segregation, and Student Friendship Networks

1.1.2.1 Desegregation and Detracking

The disparity in academic outcomes between white and minority schools during the time of official racial segregation was impossible to ignore and eventually forced the Supreme Court to mandate racial integration. Although racial desegregation was partially successful in creating more heterogeneous schools, students remained largely segregated by being assigned to different tracks based on academic achievement. Segregation mostly shifted from existing between schools to existing between classrooms within schools. Assigning students to different course sequences based on academic achievement, or tracking, has been strongly criticized for increasing inequality (Oakes et al. 1990). There has been a movement to detrack schools for this reason (Yonezawa, Wells, and Serna 2002). Although most schools in the United States are now officially detracked, the institution persists informally. The schools still generally offer the same or a very similar tiered system of course sequences. The students are not officially assigned to these course sequences, they have a choice of classes to take, but the students still sort into stratified groups (Moody 2001). There is still disagreement on the consequence of integrating the segregated informal tracks into integrated classrooms with a more heterogeneous student body. Some researchers claim that would reduce instruction inequality reducing performance for all students, others claim that it would increase performance of the low-achieving students but with a negative impact on performance amongst the high-achieving students, and finally some claim that it would reduce inequality without reducing performance of gifted students (Argys, Rees, and Brewer 1996). It may be the case that all such outcomes are possible depending on the details of how classrooms are integrated. The small-school movement claims that its approach can facilitate differentiated instruction for students of varying abilities while maintaining a cohesive, integrated community.

1.1.2.2 Small Schools and Small Learning Communities

Small school reforms advocate the creation of small learning communities (SLCs) within a larger secondary school. These SLCs are also called schools-within-a-school. The SLC is a group of students who take all their classes together with a common set of teachers. The students and teachers in the small learning community remain together as they advance from grade to grade. Advocates of small schools reform claim that SLCs have numerous advantages such as facilitating cross-class coordination between the teachers. A prominent claim is that a smaller cohort of students that has more frequent interactions develops a beneficial social network. One hypothesized benefit is that small learning communities generate social capital in the form of denser social networks which are norm enforcing (Morgan and Sorensen 1999). Another supposed benefit is that frequent and sustained interaction within the small group creates stronger ties which increase social and academic engagement. In a study of middle school students, Wentzel, Caldwell, and Barry (Wentzel, Barry, and Caldwell 2004) found that having fewer reciprocated friendship nominations was associated with lower academic achievement. In contrast to a large heterogeneous population which is housed within a single building but allowed to segregate within it, a small learning community might be able to successfully integrate students with these frequent and sustained interactions and consequently reduce inequities. Moody (2001) finds that when friendship choices are limited, for example by being in a relatively small school, segregation is reduced.

1.1.3 Student Friendship Selection

On the other hand, there is robust evidence that students tend to form friendships that are homophilous on numerous attributes including academic achievement (Schaefer, Haas, and Bishop 2012). Partially this effect follows from proximity because students of similar ability share classes either due to tracking, whether explicit or implicit. However, a significant relationship between GPA similarity and friendship formation still remains after controlling for the classes that the students share (Flashman 2012). Frank, Muller, and Mueller (2013) show that friendships are most likely to form within clusters of students taking courses together. They call these clusters local positions. Goodreau, Kitts, and Morris (2009) find that females form more friendships than males and additionally are more likely to form triangles. Schaefer, Simpkins, Vest, and Price (2011) found that shared extracurricular activities are a significant factor in friendship formation and maintenance.

All of the studies in the previous paragraph make use of data from the National Longitudinal Study of Adolescent Health (Add Health). These data have been used extensively for studies of selection and influence (Harris et al. 2009). There are two large public high schools with saturated samples in the Add Health data. Schaefer, Haas, and Bishop (2012) studied one of these schools and found significant selection and influence on smoking, as well as smoking enhanced popularity. Flashman (2012) also studied these two schools plus the six smaller schools with saturated samples. She found significant influence and selection on GPA rank in the two large schools, but the smaller schools lacked sufficient power. Although Moody (2001) did find that smaller schools in the Add Health sample which are relatively heterogeneous have lower segregation than larger schools with a similarly heterogeneous population, none of the schools in the data were identified as implementing small learning communities.

1.2 Research Questions

Proponents of small schools reform claim that SLCs promote a more cohesive social environment and that this is beneficial for norm enforcement and keeping students engaged. These claims of social cohesion would imply that the network boundary is mostly closed and that the students in the SLC mostly befriend one another rather than students outside the SLC.

Hypothesis 1 *The students in the small learning community are more likely to nominate alters in the SLC than others in the same grade but not in the SLC.*

Previous scholarship points to significant assortative mixing on academic achievement even after controlling for propinquity in the form of shared courses. Homophily on attributes associated with academic achievement is a strong force in shaping student social networks. Even if the small learning community does increase social cohesion, it is unlikely that it will be able to completely eliminate segregation on academic achievement.

Hypothesis 2 *Homophily in academic performance as measured by GPA will still play a significant role in friend selection within the SLC.*

We might expect that assortative mixing on GPA is somewhat higher for choosing study partners than friends to confide in or hang out with.

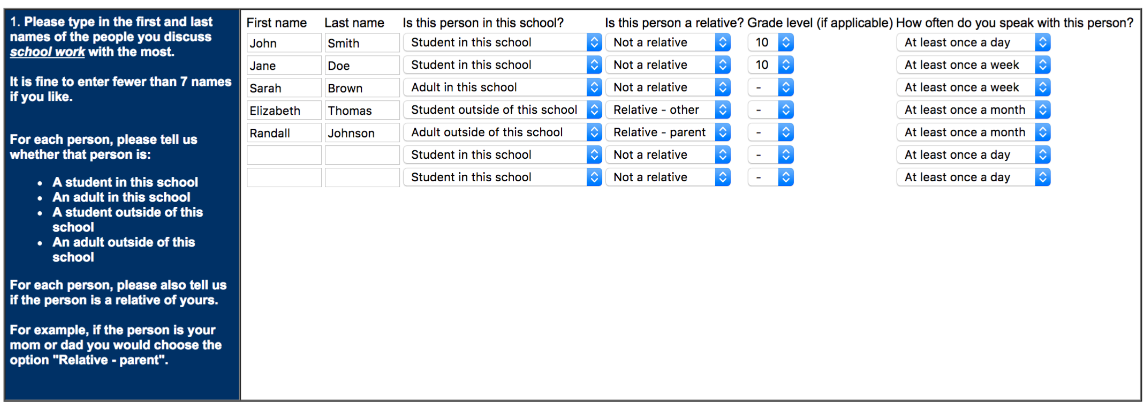
Hypothesis 3 *Academic achievement will be more important in choosing study partners (Q1) than friends to confide in (Q2) or friends to hang out with (Q3).*

1.3 Data

1.3.1 Data Summary

Social network surveys were administered at a large urban public high school in three waves at the end of the spring semesters in three consecutive years, referred to as Waves 1, 2, and 3. The surveys followed the students in one small learning community from 10th to 12th grade. It will be referred to as Small Learning Community A or SLC-A. SLC-A had a focus on math and science. Its first cohort of students (10th grade in Wave 1, 11th grade in Wave 2, 12th grade in Wave 3) is referred to as Cohort 1. In this cohort SLC-A was the only small learning community. In the first and second wave the survey was also given to a comparison group of students also in Cohort 1 but not in a small learning community. However, the comparison group in Wave 2 was not the same as in Wave 1. In Wave 3, Cohort 1 was in 12th grade and additional surveys were conducted with two small learning communities in 10th grade that year which we'll call Cohort 3. The small learning communities from Cohort 3 that were surveyed were SLC-A and another small learning community are referred to as SLC-B. SLC-B had a focus on writing.

Respondents answered multiple choice questions, name generator self-report relational questions, and relational questions about third-parties. The relational questions are described in Sections 1.3.2 and 1.3.3. The multiple choice questions are described in Section A.1. Survey were administered on a computer. Pictures of the self-report and third-party report survey instruments are shown in Figures 1 and 2.



1. Please type in the first and last names of the people you discuss *school work* with the most.

It is fine to enter fewer than 7 names if you like.

For each person, please tell us whether that person is:

- **A student in this school**
- **An adult in this school**
- **A student outside of this school**
- **An adult outside of this school**

For each person, please also tell us if the person is a relative of yours.

For example, if the person is your mom or dad you would choose the option "Relative - parent".

First name	Last name	Is this person in this school?	Is this person a relative?	Grade level (if applicable)	How often do you speak with this person?
John	Smith	Student in this school	Not a relative	10	At least once a day
Jane	Doe	Student in this school	Not a relative	10	At least once a week
Sarah	Brown	Adult in this school	Not a relative	-	At least once a week
Elizabeth	Thomas	Student outside of this school	Relative - other	-	At least once a month
Randall	Johnson	Adult outside of this school	Relative - parent	-	At least once a month
		Student in this school	Not a relative	-	At least once a day
		Student in this school	Not a relative	-	At least once a day

Figure 1. Interface for the own-tie survey

Note: This shows the input form for the first self-report question. The other three questions were shown further down on the same page.

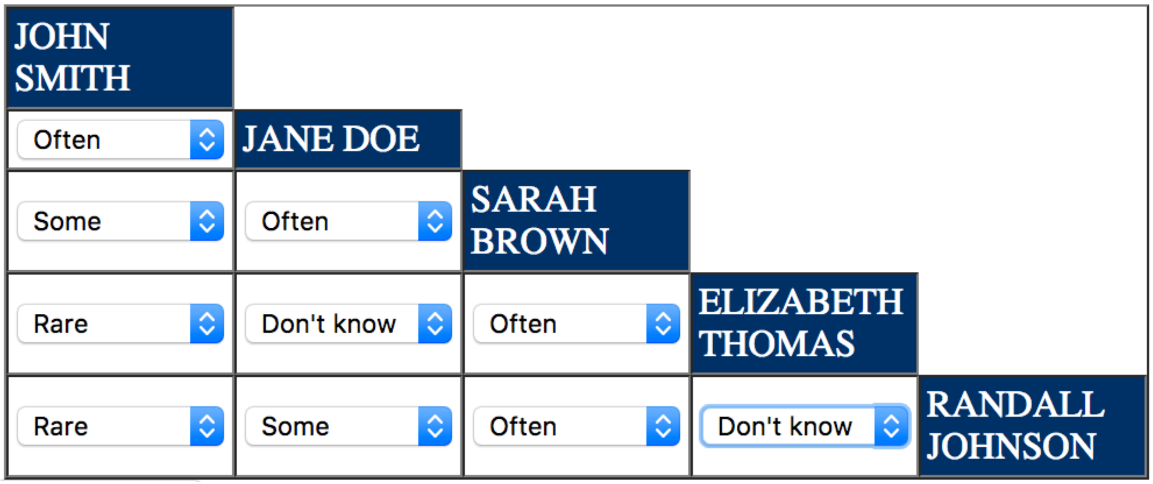


Figure 2. Interface for the third-party links survey

Note: This shows the input form for the third party links survey. The size of the input matrix would depend on the number of unique alters nominated in the self-report.

In Waves 2 and 3 the roster with student ID photos was loaded into the database and when a name was ambiguous the respondent could identify the correct person by their photograph.

In addition to the surveys, there are also paper student transcripts. These transcripts span from the fall freshman semester to the fall senior semester for our target SLC-A Cohort 1. The transcripts include student demographics such as gender and ZIP code of residence and include each semester's absences, courses, course grades, semester GPA, and cumulative GPA. Gender, ZIP code, semester GPAs, and cumulative GPAs have been input by a student worker. Not all survey respondents have matched to a transcript, so there is some missing data. In particular, transcripts are missing for most of Cohort 3.

1.3.2 Self-report

The surveys include four name generator questions asking who the student

1. discusses school with the most,
2. discusses personal and private concerns or worries with the most,
3. hangs out with the most,
4. doesn't get along with.

For each relation the student is asked to nominate up to seven alters.

There were some variations on the wording of the survey prompts across waves. Question 4 in Wave 1 asked the respondent to nominate people that the student avoids or would rather not spend time with. In Waves 2 and 3, Question 4 asked the respondent to nominate people who make it difficult to do school work. Another potentially significant difference is that questions in Wave 3 ask about relations over the entire school year but the other questions ask about the present. The exact text of each of the name generators is shown in Table 8 in Appendix A.2.1. For each nomination the respondent was asked to specify person type (student or adult; in the school or out of the school), kinship (no relation, parent, sibling, other family), and how often they speak (rarely, at least once a month, at least once a week, at least once a day).

1.3.3 Third-party Reports

In addition to self-report on these four relations, participants were asked about the relationship between the alters that they nominated in the self-report questions. This is called the third-party report or evaluation. This is a complete egocentric network census design. In contrast to the four own-tie relations that were surveyed, only one relation was measured for the third-party report: how often do these two people speak. The exact text is given in Appendix A.2.2. Additionally, the frequency categories for the own-tie and third-party report were not exactly the same. For the third-party report, speaking frequency categories were

- often,
- sometimes,
- rare,
- don't know.

1.3.4 Number of Respondents

In Waves 1, 2, and 3 there were 95, 84, and 73 respondents from SLC-A Cohort 1. There are 54 students in SLC-A Cohort 1 that responded in all three waves and 47 of these match to transcript data. In Waves 1 and 2, 79 and 88 students in the same grade but not in the small learning community were surveyed for comparison. In Wave 3, there were 69 Cohort 3 respondents in SLC-A and 79 in SLC-B.

1.3.5 Network Construction

Unless specified that network corresponds to a specific name generator question, the networks have a tie if there was a nomination from ego to alter for any of the first three network survey questions. Unless otherwise specified networks are of the SLC-A students from Cohort 1 (SLC-1). The SLC-A Cohort 1 in Waves 1 through 3 is graphed in Figures 4, 5, and 6.

1.3.6 Network Motifs

Network motifs are subgraphs which occur in a network more often than would be expected by chance. Examining the occurrence the local structure of networks with motifs can test theoretical claims about tie formation and may inform subsequent statistical modeling (Wasserman and Faust 1994; De Nooy, Mrvar, and Batagelj 2018). The simplest motifs involve two or three nodes – dyads and triads. The counts of the dyads and triads within the possible isomorphism classes are called dyad and triad censuses. Dyads may be in one of three isomorphism classes: mutual, asymmetric, and null. The sixteen triad isomorphism classes are labeled according to the scheme introduced by Holland and Leinhardt (1970) with three numbers counting the Mutual, Asymmetric, and Null dyads (the MAN labeling) and a possible fourth character – D for down, U for up, C for cyclic, and T for transitive.

Figure 3 plots the z-scores for the triad census testing against a conditional random graph null hypothesis, conditioning on the dyad census (the $U|MAN$ distribution). The social network of the SLC-1 students has significant triadic patterning not explained by the lower order graph features (the dyad census), as is the case in a majority of social networks (Faust 2010). Plots of the triad census z-scores for the networks corresponding to the individual name generator questions are shown in Appendix B.1. Overall the three name generators appear to have a similar pattern in the triad census. Forbidden triads 021C, 11D, 11U, and 201 all occur less than expected by chance. The triads 030T, 120D, 120U, 210, and 300 all occur significantly more than expeced by chance. This is roughly consistent with a hierarchical clusters balance-theoretic model of tie formation (De Nooy, Mrvar, and Batagelj 2018).

Density, reciprocity, and transitivity scores are shown in Tables 1, 2, 3. Reciprocity

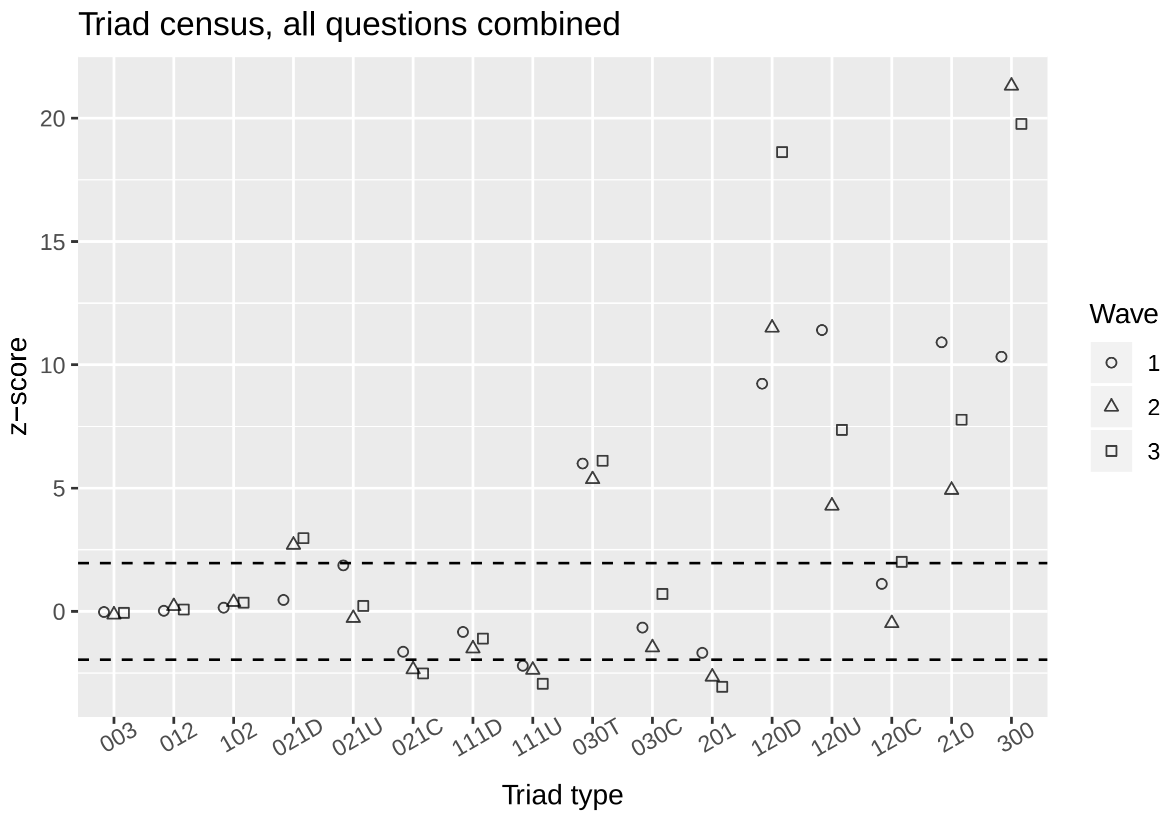


Figure 3. Triad census of the SLC-1 all questions combined networks

is calculated using Garlaschelli and Loffredo's definition as the correlation between mutual links (2004). For each of the observed networks, I simulated 1000 draws from a density-conditioned random graph distribution and the observed reciprocity was not exceeded. Similarly, I simulated 1000 draws from a dyad census-conditioned random graph distribution for each of the observed networks and the observed transitivity was not exceeded. The level of transitivity goes up each wave but reciprocity does not show this pattern. This is consistent with the findings of Doreian et al. (1996) that transitivity has a longer time scale than reciprocity.

Table 1. Density each network type (the three name generators Q1, Q2, Q3 plus the collapsed network with all questions combined QC) across all three waves

	Wave 1	Wave 2	Wave 3
Q1	0.03	0.05	0.05
Q2	0.02	0.03	0.03
Q3	0.03	0.04	0.04
QC	0.04	0.06	0.06

Table 2. Reciprocity scores for each network type (the three name generators Q1, Q2, Q3 plus the collapsed network with all questions combined QC) across all three waves

	Wave 1	Wave 2	Wave 3
Q1	0.53	0.50	0.56
Q2	0.60	0.49	0.54
Q3	0.56	0.53	0.52
QC	0.58	0.52	0.55

Table 3. Transitivity scores for each network type (the three name generators Q1, Q2, Q3 plus the collapsed network with all questions combined QC) across all three waves

	Wave 1	Wave 2	Wave 3
Q1	0.33	0.37	0.42
Q2	0.23	0.28	0.49
Q3	0.38	0.42	0.52
QC	0.35	0.35	0.41

1.4 Methods

1.4.1 Modeling Approach

1.4.1.1 Challenges

There is a large body of literature that studies how social networks impact a variety of academic behaviors and outcomes including course-taking, grades, absences, and persistence. Much of this literature focuses only on a single network component and utilizes a single cross-sectional network measurement. Even with longitudinal network data, it remains difficult to disentangle interdependent network processes. The social network within a school influences academic achievement and academic achievement impacts network changes over time, resulting in complex dynamics. In fact it is not possible in general to distinguish selection and influence (also called homophily and contagion) in observational social network data (Shalizi and Thomas 2011). Furthermore, there are a variety of features of network structure and composition that can impact academic outcomes such as the attitudes and behaviors of peers (influence, contagion), access to resources through peers (information, helping), the density and quality of social ties (social engagement, norm-enforcement), and structural position (brokerage, power).

The focus of this project is on the friendship network rather than academic outcomes. Even setting aside academic outcomes, the study of friendship selection is still plagued by interdependence. Students would be more likely to become friends if they encounter one another often (propinquity), but also if they have shared interests. However, shared interests can result in propinquity, such as by joining the same club. Students can also have a preference for friends that are like them (homophily, assortative mixing) on a variety of traits such as race, gender, and socio-economic background. A student's choice in friends is not independent of the others' choices, the student may be likely to return a friendship overture (reciprocity) or to befriend friends of friends (triadic closure). Again, all these interact, assortative mixing can increase triadic closure and triad closure can amplify assortative mixing (Goodreau, Kitts, and Morris 2009).

These complications mean that simplifying assumptions must be made if we are to try to model. For example, stochastic actor-based models (SABM) attempt to distinguish the selection and influence by making a distributional assumption (generalized linear model with exponential link) and by assuming that observable characteristics carry all of the dependence between latent traits and observed outcomes (Shalizi and Thomas 2011; Snijders, Van de Bunt, and Steglich 2010). If this latter assumption is violated, latent homophily could result in apparent contagion where there is none. For this reason, multiple approaches are applied and results compared.

1.4.1.2 Network Logistic Regression

A network logistic regression models the conditional probability of a directed or undirected tie. The logistic regression is a generalized linear model that models a probability with a Bernoulli distribution and a log-odds link function, thus

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta X \tag{1.1}$$

where p is tie probability, X are the predictors, and β are the coefficients.

This model makes the assumption that the ties are independent. An important factor in getting unbiased estimates is to include as many of the relevant covariates as possible. For this study, unfortunately some important covariates like race, course enrollments, and activity participation are missing. There is longitudinal data. One way to help mitigate the issue of tie interdependence is to condition on network measurements from previous time periods by including whether there was a tie in either direction, the number of friends that were shared in common, and other network features. Another helpful approach is to add mixed effects on sender and receiver to capture individual sociality and popularity in the current time period.

1.4.1.3 Exponential Random Graph Models

Exponential random graph models (ERGMs) allow us to predict the joint probability of ties in a network. ERGMs can directly model network interdependencies including edge dependencies such as reciprocity and dyad dependencies such as triad closure.

For a random graph Y the probability may be written as

$$P(Y) = \frac{\exp(\theta^T z(Y))}{c(\theta)} \tag{1.2}$$

where $z(Y)$ is a vector of sufficient statistics, θ is a vector of model parameters, and $c(\theta)$ is the normalizing constant.

Temporal ERGMs can model network evolution over time by specifying different functions for the creation and dissolution of ties. A simple way to model this for two waves is to condition on the previous network and include interaction effects.

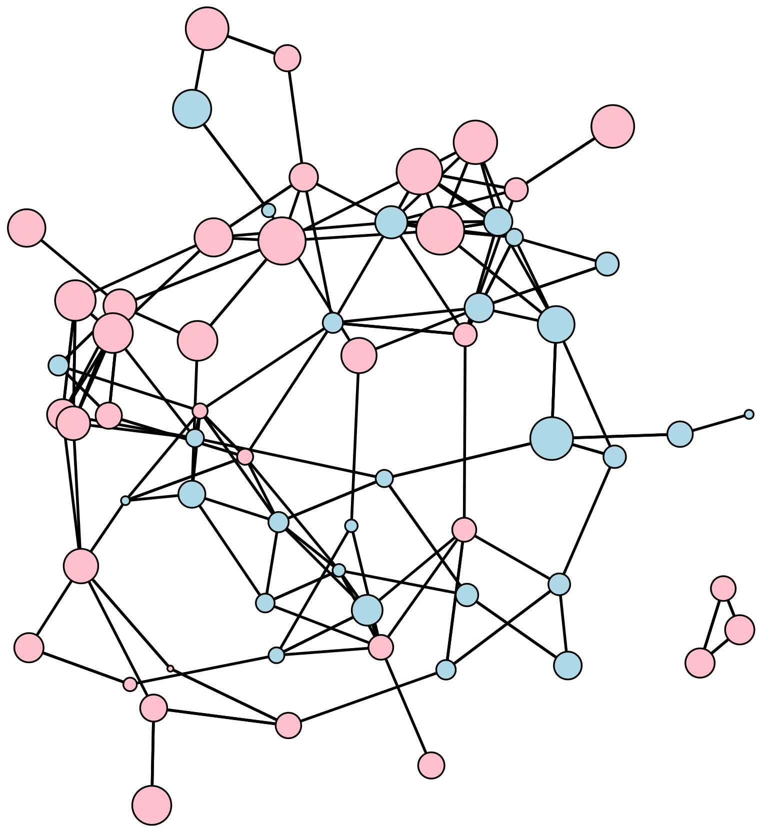


Figure 4. Small learning community network in Wave 1

Note: Node color indicates gender - blue is male, red is female. Node size indicates GPA.

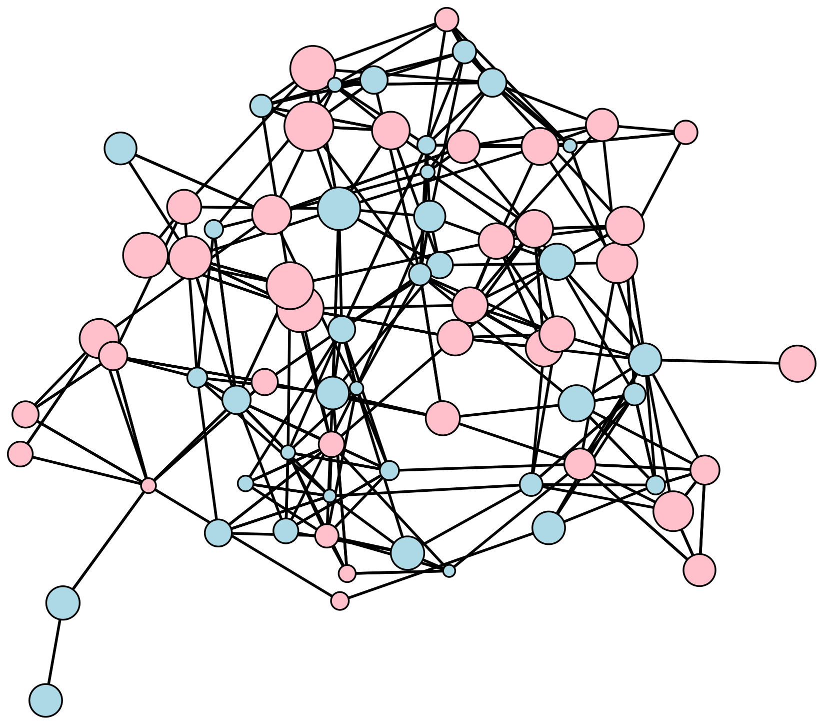


Figure 5. Small learning community network in Wave 2

Note: Node color indicates gender - blue is male, red is female. Node size indicates GPA.

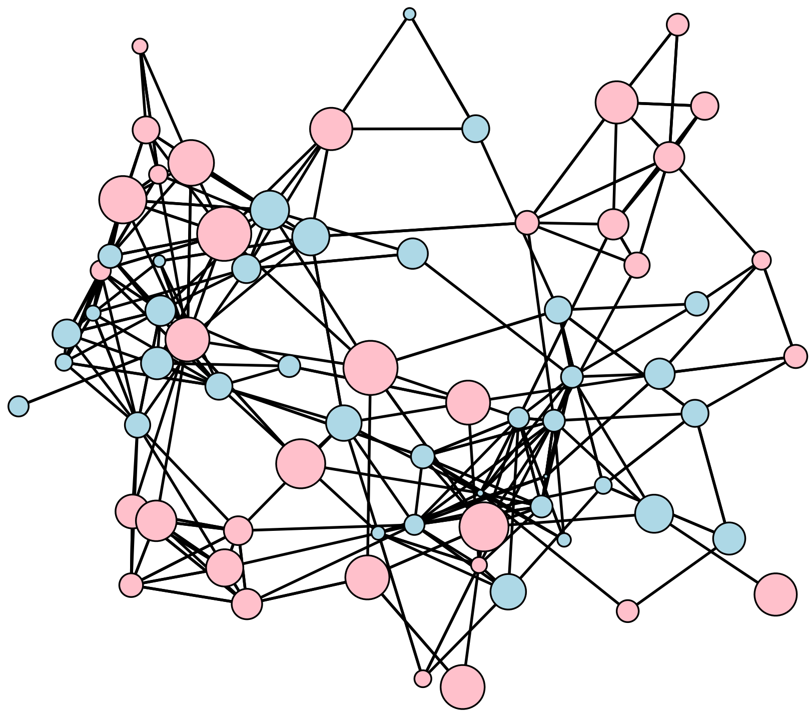


Figure 6. Small learning community network in Wave 3

Note: Node color indicates gender - blue is male, red is female. Node size indicates GPA.

1.4.2 Model Implementation

1.4.2.1 Cross-sectional Exponential Random Graph Models

Cross-sectional ERGMs are fit on the SLC-1 network in Waves 1 through 3. For Waves 1 and 2, I also fit models on the network which includes the control group of students in the same grade who are not in the small learning community. I include terms for the number of edges in the graph, the geometrically weighted out-degree, for the sociality of females, high GPA, and of SLC-1 students; for reciprocity in general and for reciprocity amongst SLC-1 students and amongst students not in SLC-1; for selective mixing on ZIP code, on gender, on GPA similarity, and on being in the SLC or not. I also include terms for mutuality and transitive closure.

1.4.2.2 Longitudinal Logistic Regression on Network Data

I fit a network logistic regression conditioned on the network in the previous wave. The lagged network features included are the presence or absence of an outgoing tie from ego to alter, an incoming tie from alter to ego, and the number of 2-paths from ego to alter (transitive potential). Terms are included for homophily on ZIP code, gender, and GPA. GPA similarity is given by

$$GPA_{sim} = 1 - \frac{\mid GPA_1 - GPA_2 \mid}{\max GPA_i - \min GPA_i} \tag{1.3}$$

Hypothesis testing is performed with the quadratic assignment procedure (QAP) permutation test using Dekker's "semi-partialling plus" procedure.

1.4.2.3 Longitudinal Exponential Random Graph Models

Longitudinal ERGMs are fit for Waves 2 and 3. Terms are added for the number of edges, for the geometrically weighted out-degree, and for homophily on ZIP code, gender, and GPA. I also condition on the network in the previous wave with lagged terms for the presence of an outgoing tie from ego to alter, for the transitive potential (number of 2-paths from ego to alter), and I add interaction of the lagged tie with the lagged transitive potential and with GPA similarity.

1.5 Results

1.5.1 Cross-sectional ERGM

Table 4 presents the results of the cross-sectional ERGMs for each wave. There does not appear to be evidence for higher sociality amongst females in this dataset, as Goodreau et. al (2009) find in AddHealth, nor evidence of sociality based on GPA or being in the SLC. Estimates of the selective mixing coefficients on ZIP code are larger with the expanded network boundary, suggesting that the cohesiveness of the SLC may reduce assortativity on this trait. Selective mixing on GPA is observed as predicted by Hypothesis 2. In both the models of the network with the control group included for Waves 1 and 2, the coefficient for reciprocity amongst students not in SLC-1 is higher than for reciprocity amongst student within SLC-1. However, all of these coefficients are negative. This suggests that reciprocity is higher across groups than within. There does not appear to be a significant difference in transitive closure

	SLC-1 W1	SLC-1 W2	SLC-1 W3	Full W1	Full W2
Sociality					
Edges	-5.87^{***}	-5.22^{***}	-5.54^{***}	-9.07^{***}	-7.49^{***}
	(0.38)	(0.24)	(0.28)	(0.59)	(0.42)
Out-degree	2.75^{**}	1.83^{**}	2.83^{*}	1.48^{***}	1.12^{**}
	(0.83)	(0.70)	(1.15)	(0.39)	(0.37)
Female	-0.02	-0.01	0.01	0.01	-0.02
	(0.07)	(0.05)	(0.04)	(0.06)	(0.04)
GPA	0.06	0.00	0.02	0.03	0.01
	(0.04)	(0.03)	(0.02)	(0.03)	(0.03)
SLC-A				0.33^{**}	0.16
				(0.12)	(0.11)
Selective Mixing					
Same ZIP	0.37^{**}	0.37^{***}	0.47^{***}	0.43^{***}	0.47^{***}
	(0.11)	(0.08)	(0.08)	(0.10)	(0.08)
Same gender	0.19	$0.15^{\cdot}$	0.05	0.31^{**}	0.18^{*}
	(0.13)	(0.09)	(0.10)	(0.11)	(0.08)
GPA similarity	0.91^{**}	0.55^{*}	0.69^{**}	0.90^{**}	$0.48^{\cdot}$
	(0.31)	(0.23)	(0.25)	(0.32)	(0.27)
Both SLC or Both not SLC				2.57^{***}	1.97^{***}
				(0.47)	(0.29)
Mutuality					
Mutuality	3.67^{***}	2.89^{***}	2.97^{***}	6.13^{***}	5.60^{***}
	(0.29)	(0.22)	(0.23)	(1.34)	(0.73)
Not in SLC				-1.27	-1.95^{*}
				(1.41)	(0.83)
Within SLC				$-2.54^{\cdot}$	-2.67^{***}
				(1.37)	(0.76)
Transitive Closure					
Transitive triads	0.97^{***}	1.23^{***}	1.25^{***}	1.07^{***}	1.16^{***}
	(0.12)	(0.11)	(0.12)	(0.10)	(0.10)
AIC	1152.06	1866.60	1730.19	1812.98	2613.62
BIC	1208.49	1925.95	1788.54	1910.69	2711.98
Log Likelihood	-567.03	-924.30	-856.10	-893.49	-1293.81

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 4. Cross-sectional exponential random graph models

Note: For all three waves there is a cross-sectional ERGM on just SLC-1 students. For the first two waves (W1, W2) there is a full model with both the SLC-1 students and the control group. Standard errors in parentheses.

when including the control group, as might be expected if students within the SLC were more or less likely to complete transitivity.

1.5.2 Longitudinal Network Logistic Regression

Terms for sociality on gender and GPA were not even weakly significant so are not included. Results are displayed in Table 5. In Wave 2 the coefficient for GPA similarity is positive and significant for all questions. For the combined question network and for questions 1 and 3, it is the largest positive coefficient. For question 2, only the coefficient on the lagged tie is larger. This provides support for Hypothesis 2. For both waves, the coefficients for GPA similarity are larger for question 1 than for questions 2 or 3. The QAP procedure does not produce standard errors so we cannot test if the coefficients are significantly different, but the direction is in agreement with Hypothesis 3. For all questions the coefficient estimates for assortative mixing on GPA similarity and ZIP code are larger in Wave 2 than in Wave 3. This might be an indication that the SLC is having some success in integrating the students.

1.5.3 Longitudinal ERGM

Table 6 presents the results of the longitudinal ERGMs for Waves 2 and 3. The positive coefficients for GPA similarity in Wave 2 are in line with Hypothesis 2. The coefficients for the interaction of GPA similarity with the lagged tie are not significant so we would accept the null that GPA similarity is equally important for creating and maintaining friendship ties. The drop in the coefficient for GPA similarity from Wave 2 to 3 may indicate that the SLC is having some success in integrating the students over

	Q1 W2	Q1 W3	Q2 W2	Q2 W3	Q3 W2	Q3 W3	QC W2	QC W3
Intercept	-6.26^{***}	-5.71^{***}	-5.84^{***}	-5.65^{***}	-6.04^{***}	-3.77^{***}	-5.96^{***}	-4.58^{***}
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Lag ego $\rightarrow$ alter tie	2.28^{***}	2.25^{***}	3.69^{***}	3.07^{*}	2.50^{***}	2.04^{***}	2.35^{***}	2.22^{***}
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)
Lag alter $\rightarrow$ ego tie	$1.20^{\cdot}$	2.28^{***}	0.41	2.16^{*}	1.75^{***}	1.34^{*}	1.28^{*}	1.40^{***}
	(0.07)	(0.00)	(0.67)	(0.01)	(0.00)	(0.01)	(0.01)	(0.00)
Lag transitive potential	0.53	0.84^{***}	0.60	1.90^{***}	0.78^{*}	1.19^{***}	$0.59^{\cdot}$	0.69^{***}
	(0.30)	(0.00)	(0.57)	(0.00)	(0.03)	(0.00)	(0.06)	(0.00)
Same ZIP	$0.67^{\cdot}$	-0.43	0.75	-0.46	0.24	-0.86^{*}	$0.47^{\cdot}$	-0.27
	(0.08)	(0.10)	(0.16)	(0.34)	(0.43)	(0.02)	(0.06)	(0.26)
Same gender	0.07	0.88^{***}	0.08	0.87^{*}	$0.53^{\cdot}$	0.22	$0.43^{\cdot}$	0.62^{***}
	(0.83)	(0.00)	(0.80)	(0.03)	(0.07)	(0.45)	(0.06)	(0.00)
GPA similarity	3.42^{***}	1.81^{*}	2.61^{***}	0.41	2.56^{***}	-0.10	3.00^{***}	$0.92^{\cdot}$
	(0.00)	(0.01)	(0.00)	(0.77)	(0.00)	(0.93)	(0.00)	(0.09)
AIC	589.37	749.55	323.83	301.42	592.73	557.40	1049.28	1241.62
BIC	627.86	791.52	358.99	339.91	632.49	596.85	1091.26	1285.29
Deviance	575.37	735.55	309.83	287.42	578.73	543.40	1035.28	1227.62
Null deviance	2503.65	4117.29	1555.42	2503.65	2997.17	2869.63	4117.29	5242.97
Num. obs.	1806.00	2970.00	1122.00	1806.00	2162.00	2070.00	2970.00	3782.00

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 5. Longitudinal network logistic regressions

Note: I construct a regression for each question and for all questions combined (Q1,Q2,Q3,QC) in Waves 2 and 3 (W2, W3) conditioned on the network in the previous wave. Transitive potential is the number of dyad-wise shared partners - the number of 2-paths from ego to alter. Hypothesis tests are performed with QAP. The value reported in parentheses is the proportion of draws from the QAP for which the absolute value of the null coefficient is greater than the observed estimate.

	W2 | W1	W3 | W2
Edges	-5.55 (0.47)***	-4.58 (0.43)***
Out-degree	-0.37 (0.59)	0.62 (0.73)
Lag ego $\to$ alter tie	2.26 (1.01)*	2.54 (0.83)**
Lag transitive potential	0.89 (0.19)***	0.86 (0.13)***
Lag transitive potential $\times$		
Lag ego $\to$ alter tie	-0.50 (0.26)$^{\cdot}$	-0.22 (0.20)
Same gender	0.42 (0.19)*	0.67 (0.17)***
Same ZIP	0.45 (0.19)*	-0.24 (0.19)
GPA similarity	2.53 (0.58)***	0.86 (0.54)
GPA similarity $\times$		
Lag ego $\to$ alter tie	1.34 (1.22)	0.63 (1.04)
AIC	1067.29	1278.78
BIC	1121.26	1334.92
Log Likelihood	-524.65	-630.39

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 6. Longitudinal exponential random graph models

Note: Standard errors in parentheses.

time. It is curious that the coefficient on ZIP code homophily goes up from Wave 2 to 3 in contrast to the decline in the cross-sectional ERGMs, but the difference is not significant.

Additional results are reported in Appendix B.

1.6 Conclusion

The results suggest that this small school reform was not able to eliminate segregation of social ties on academic achievement. It is not known if the teachers did any ability grouping within their classrooms. However, grouping the students together into the same classroom was not enough on its own to provide full integration, much as grouping students together into the same school was not enough when the students sorted into different tracks. More work is needed to better understand the impact of small-schools reform on student friendship selection.

This work replicates the findings of Goodreau et al. (2009) that selective mixing and transitive closure structure the process of student friendship formation. The cohesion of the SLC was not sufficient to eliminate these structures. On the other hand, homophily on GPA appears to go down over time. This might be an indication of the success of the SLC at integrating the students.

A small learning community is somewhat like a local position as described in Frank et al. (2013) in that the students take all of their classes together, only this local position is larger. Students in the SLC are also most likely to nominate other students within the small learning community, similar to Frank's finding that students are most likely to nominate others in their local position. Grouping the students together in the SLC does appear to mostly limit their friendship choices to other students in the SLC. Based on Moody's findings this would suggest that segregation will at least be lower in the SLC although still not yet eliminated.

The longitudinal ERGM model from Section 1.4.2.3 is close to the specification found in Schaefer et al. (2011), but they use a different measure of GPA similarity. In that model, Schaefer et al. find evidence of selection on GPA in one of the two large schools. Across the models explored thus far GPA has a relatively strong impact, as Flashman (2012) found in her study. Every paper mentioned so far in this conclusion is based off the AddHealth data set. This work makes a contribution by replicating key findings with a somewhat more recent dataset. Furthermore the study is set in the context of a small-school reform while there continues to be debate about the effectiveness of this reform.

Chapter 2

PROJECT 2: INFORMANT COMPETENCE

2.1 Introduction

This project explores the role of third-party informants in network measurement. Better understanding of the relationship between third-party and self-reports of social relations may help us to more accurately and efficiently measure social networks. Third party informants can be essential to deal with non-response and useful to improve data quality. The most common way that third party reports are collected in a network census is by asking respondents to report on all dyads in the population. This measurement design is called the cognitive social structure and is abbreviated as CSS (Krackhardt 1987). Its use is limited to small populations such as a single classroom because the number of dyads grows quadratically with the number of individuals N. Survey burden may be reduced by asking respondents only about a random subsample of dyads instead of all (Butts 2003), but such an approach still reaches a limit when respondents are no longer well acquianted with the entire population. These observations motivate research that would help to optimize decisions regarding which informants are asked about which relationships.

Using the data on social ties among the students in a small learning community described in Section 1.3, I develop an informant competence model using mixed-effects ordinal location scale regression. This unique ego network census data and heteroskadastic modeling approach permit inquiry into the relationship between individual characteristics and network position of the informant and competence in a popluation that is too large for a cognitive social structure survey. I find that informant competence corresponds positively with academic achievement and that girls are better informants than boys. Furthermore, I find that informant competence corresponds positively with measures of the informant's social proximity to the dyad in question as well as with measures of the informant's network centrality.

These findings have at least two implications for network surveys. First, the salience of social proximity highlights the inherent tradeoff with using designs that use a balanced random assignment of dyads to informants in large networks. While such designs have the important benefit of reducing the potential bias of missing reports, they may also include many reports that contain very little information about a relation — i.e., the larger the network, the greater the number of reports from informants that are not in sufficient social proximity to observe them. Second, in the case of measuring a network with a small number of informants, rather than screening prospective informants only on likely global competence as suggested in (Marsden 2005), the set of informants could be selected to optimize coverage such that all dyads are evaluated by informants likely to be knowledgeable about their relationship. This line of research remains largely untapped and may yield substantial benefits for studies relying on network census.

2.1.1 Sampling Method

As with any survey method, when conducting a network survey we must define the population of interest and how we will sample from this population. The importance of getting the population definition right for the theoretical question of interest is often more important than with traditional individual focused methods because with network surveys the omission of key actors or inclusion of non-salient actors may have a large impact on the overall structure of the network.

Social groups are rarely if ever closed systems, so it is important to have a theoretical basis for defining the cutoff between the individuals that need to be included in the analysis and those whose impact on the social process of interest may be safely ignored. This is termed defining the network boundary. In Butts' review of social network methods (2008a), he defines three manners in which the network boundary may be defined: exogenously, endogenously, methodologically. An exogenous network boundary definition is driven by theory before having made any network measurement. Who do we expect the relevant players to be in this social process? If we are interested in the adoption of technology in the classroom, then it is theoretically justified to focus on the network of teachers and school administrators if we do not expect students to play an important role in that process. An endogenous network boundary definition starts with a core set of nodes and expands outward from this core by link-tracing until the nodes being sampled are sufficiently removed that their influence may be considered negligible. The network boundary may be defined by the methodology. This is most often the case with archival data on social interactions or affiliations where the boundary is defined by participation in the system which generates that data, such as the use of an online social network. Both with endogenous and methodological definitions, it is still important to evaluate critically whether the network boundary is theoretically justifiable.

With the network boundary defined, it must then be decided how the population is to be sampled. For small populations it is preferable to survey the entire network. This is called a network census. This gives a view of each actor's position within the whole network structure which may be of theoretical importance. If the population is too large for a network census, then individuals may be sampled by random selection. In this case we do not get a complete view of the network and the actors' positions in it, but that may be unnecessary if the theoretical focus is on local interactions. A third category of sampling methods are link-trace methods such as snowball sampling. Link-trace sampling begins with an initial seed set of individuals to be surveyed and each subsequent wave is selected from the set of alters that have been nominated but not yet surveyed. Link-trace sampling can be useful for surveying hard to reach populations. It may also be useful to permit inference of some whole network properties when a complete network survey is not feasible.

2.1.2 Survey Method

2.1.2.1 Self-report

There are several methods by which social relations may be measured. The most common method is self-report where individuals are surveyed about their own relationships. These are also called own-tie reports. There are problems with using only own-tie reports for network measurement. It is desirable to have complete or nearly complete response because missing information on relations on only a small number of actors can have a relatively large impact on network statistics. A small percentage of non-response can results in a larger percentage of dyads for which there is missing data. Achieving high response rates can be difficult in school settings due to the need for parental consent forms. Measurement error is a problem with own-tie reports due to only having one or two reports on each dyad.

Self-report network relations may be surveyed with different instruments, such as rosters and name generators. With name generators the respondent is asked to list the names of alters with whom they have a given social relation. An advantage of name generators is that they can allow open-ended nomination that is not restricted to a predefined population. This may be necessary if the population isn't known beforehand. If the population is known, this advantage can still be beneficial to allow the researcher to test the appropriateness of their network boundary definition.

In the National Longitudinal Study of Adolescent Health (Add Health) study students were asked to nominate up to 5 male friends and 5 female friends. A drawback of the name generator is that it may be difficult to uniquely identify alters. If information on the population population is available, this problem can be dealt with by having the respondent confirm the identity of their nominations against the roster. This was done in the Add Health study. Another drawback of the name generator is that the respondent may not recall all relevant alters or their response may be truncated if they can only list up to a certain number of alters. One way to help address the recall problem is to prompt with groups that potential alters may be categorized into. The Add Health study does this in a way by prompting separately for male and female friends. There was the issue of truncation in the Add Health by limiting the number of nominations to five friends of each gender.

An alternative survey instrument to name generators is the use of a roster. In this case, the respondent is given a list of all members of the population and asked to select those with whom they have a given relation. A roster can be combined with a name generator where the respondent is invited to list names of additional alters that are not in the roster. The roster has the advantage of avoiding measurement error due to recall errors and truncation. However, use of a roster may become cumbersome or impractical when the population is too large.

2.1.2.2 Third-party Report

With third-party reports, respondents are asked about the relationship between pairs of other actors. Third-party reports can be useful for dealing with measurement error and missing data. They might also be of interest in their own right if the theoretical question relates to how relationships are perceived. It can be feasible in small networks for respondents to evaluate all possible ties. This is called the cognitive social structure. However, as the number of ties for each relation grows quadratically with the number of actors, there may be too many for respondents to evaluate them all. Butts (2003) proposes a balanced sampling method to achieve a desired number of observations per tie which he calls K-replication balanced arc sampling. A more common method of sampling third-party ties is the ego network approach. With this method, respondents are asked about the relationship between all pairs amongst the alter that they nominated by self-report. The ego network method is most commonly used in large networks that are sampled by random selection. It can be combined with a network census, but this is uncommon.

There are also multiple survey instruments for third-party reports. Generally, a roster is used in some way with third-party reports due to problems of recall associated with name generators. In a cognitive social structure approach, the roster includes all actors in the population. In an ego network approach, the third-party roster includes only the actors that the respondent nominated in self-report. When asking the respondent to evaluate the third-party relationships, this can be presented all at once with a triangular matrix listing all actors in the roster on the rows and columns. The respondent then evaluates the relationship of the between the actors of the corresponding row and column. Alternatively, each actor in the roster can be presented in succession along with the rest of the roster as a list so the respondent can evaluate all relationships of the focal actor. Watling Neal (2008) used this method in a classroom context and also included a name generator. For each student X in the class the survey included a page which asked "Please CIRCLE the names of all of the kids in your classroom that X hangs out with often" followed by a list of all the other students. Beneath this roster was the name generator question: "Are there any other kids at your school that X hangs out with often?"

2.1.2.3 Observation and Archival Data

Lastly, behavioral relations may also be measured by observation and via archival data. Observation can be done in person or by audio/video recording. This is useful for measuring observable behaviors of theoretical interest. As an example, Ridgers et al. (2011) observed children during school recess and recorded instances of verbal and physical antisocial behavior. Depending on the construct, a potential drawback with qualitative coding of observations can be limited reliability. It may be overly costly to observe and code a sufficient portion of the relevant social interactions. Observational network measurement sees limited use for this reason. Conversely, network information from archival data is seeing increased usage as such data is generated by more and more systems and is relatively easily to acquire. Archival data does not have the reliability issues that may occur with qualitative coding and it is more likely to have comprehensive coverage of the type of interaction it measures. However, a drawback of archival data is that it is generated by and for a system which serves some purpose other than answering the research question at hand. In this respect it may only give a glimpse of the underlying social processes that are of interest. A good example is a workplace email system. Using this data to study workplace interaction has the advantage that it is easy to collect and quantify *all* emails between workers, but this measure is blind to other modes of interaction such as verbal which might be more important. Furthermore, without interviewing the workers themselves use of archival data also runs a higher risk of misinterpreting its social meaning.

2.1.3 Method Validity

The conceptualization and operationalization of social ties for a network study should be driven by the theory and research question (Marsden and Campbell 1984, 2012), but researchers must also consider the validity of their measures. In this section I review literature on the accuracy of informant reports of social ties. Where informant accuracy is measured against the consensus of all informants, it will be termed competence (Romney, Weller, and Batchelder 1986).

2.1.3.1 Informant Accuracy

In a seminal series of five papers published from 1976 to 1982, Bernard, Killworth, and Sailer conducted seven experiments evaluating the accuracy of informant reports on social interactions against observational or archival data treated as ground truth (Killworth and Bernard 1976; Bernard and Killworth 1977; Killworth and Bernard 1979; Bernard, Killworth, and Sailer 1979, 1982). These have been referred to as the BKS papers. There were four experiments where archival data was collected: two with deaf teletype users (TTY1, TTY2), a third with ham radio operators (Ham), and fourth with scientists using an early email system (Email). Three experiments employed behavioral observation in: a small social science research firm (Office), a graduate program in technology education (Tech), and a college fraternity (Frat). Together with Kronenfeld, BKS review these and other studies of informant accuracy (1984). They state in their summary of the BKS papers

1. "asking people 'who do you like?' produces about the same answers as asking them 'who do you talk to?'",
2. "asking people about the significance or importance of their interactions with others is of little use since it produces no better results than simply asking them simply who they talked to", and
3. "questions about interactions two weeks previously were the least accurately reported, with better accuracy for shorter and longer times ago."

One source of error is memory decay over time. Their review of work by Sudman, Bradburn, and others found that data collection and interviewing techniques designed to account for memory decay can increase accuracy up to 10% (Sudman and Bradburn 1973, 1974, 1983). In reviewing work by D'Andrade and other they also found that errors in recall are not random and are subject to systematic distortion from cultural expectations and other factors (D'Andrade 1965, 1973, 1974). Nevertheless, based on the seven BKS experiments they concluded

1. "what people say about their communications bears no useful resemblance to their behavior",
2. "individual differences in accuracy could not be accounted for by any of the usual characteristics of people or groups, such as sex, age, time in group, centrality in group, etc.", and

3. “the error is so great that statistical and numerical techniques for washing data collected by recall instruments cannot solve the problem”.

These strong conclusions provoked a debate that included reanalysis of the BKS data and new experiments. These studies aimed to address the questions of when informant reports are accurate enough to be useful, what individual characteristics can explain differences in informant accuracy, and what techniques can be used to improve the quality of data from informant reports.

Kashy and Kenny (1990) reanalyzed the Frat, Office, Tech, and Ham data. They found that popularity in the cognitive network was strongly correlated with popularity in the behavioral network. They argue that dyadic correlations should be controlled for actor effects and partner effects. They found little correspondence between the number of outgoing ties in the cognitive network and local density in the behavior network indicating that differing thresholds of respondents is a source of inaccuracy. They also argue that the observational data may be more erorr prone than BKS acknowledge. Burt and Bittner (1981) respond to BKS IV (Bernard, Killworth, and Sailer 1979) and take issue with the fact that BKS did not present evidence of the adequacy of proposed subgroups. Their test statistics show that the TTY network is not differentiated in terms of network subgroups. They say that the network is best represented by a core-periphery structure, not as a number of cliques, and that therefore to compare the cliques between the cognitive and behavioral networks does not produce useful information. Romney and Faust (1982) examine the Tech dataset from BKS and find that the cognitive data can be used to predict structural aspects of the observed data. They find that the more similarly two people rank others, the more they interact with each other. Romney and Weller (1984) examine the Frat, Ham, Tech, Office datasets and find that they can account for much of the variance of informant accuracy by how much their individual rankings correspond to the aggregate behavior. [1]

Freeman, Romney, and Freeman (1987; 1987) respond to BKS with a study of their own where they examine recollections of attendence at a colloquium. They find that errors are not random but are biased towards long-term regularities. Regular attendees are more likely to make errors of commission by mistakenly claiming that people who regularly go to the series attended a particular event and new or sporadic attendees make more errors of omission. They argue that regular attendees are therefore better informants on long-range stable patterns but that sporadic attendees can be used to more accurately inform on a particular event. An interesting analysis in this paper examines the proximity of names in the lists generated by informants because memory research shows that in free recall lists related concepts tend to be grouped together. A multidimensional scaling plot of of the distances between the names generated by in-group informants shows a clear separation between the in-group and out-group, but in the plot for the data from the out-group there is no pattern. These findings are important because researchers are more often interested in the social structure of groups, their "enduring patterns of interpersonal relations", rather than in a particular events or exact behavior over short time periods. In that case, a bias towards the social structure is actually an advantage of recall data relative to observation. Freeman

[1] Romney and Weller (1984) interpreted the cognitive questions very differently than BKS: "There is no way to really reconstruct exactly the meaning of the recall questions to the subjects. Our model assumes that the subjects were reporting on the overall amount of communication engaged in by others (not how much the subject communicated with each person). The model assumes that each subject is observing a sample of some objective reality and making their judgement on the basis of their sample observations. To the extent that this assumption is wrong then the results will fail to reach significance. There is no way to prove that our interpretation is correct other than to look at the final results." I was surprised by this viewpoint and that BKS did not address the disagreement in their (1984) review. I agree with the original BKS interpretation of the question and think a possible alternative explanation could be that those whose individual rankings correspond to the aggregate behavior are actually more central and that it is this centrality that is linked to higher informant accuracy.

(1992) reviews research on the accuracy of reports about social affiliation and finds that although affiliation is not categorical, humans percieve it as such. However, this is not necessarily a problem because "people's mental assignments of individuals to groups correspond quite closely with observed interaction frequencies."

Given the difficulty of collecting observational data of social interactions and the limited scope of most archival data there has not been a lot of subsequent work to compare informant reports on social ties versus such ground truth. Instead there has been work to compare third-party informant reports to self-reports (informant accuracy, used differently than BKS) or to compare against other informants (competence).

2.1.3.2 Accuracy and Network Position

Many studies have looked how the network position of actors relates to their overall accuracy (individual level accuracy). Centrality is the most prominently researched correlate and is usually found to be positive correlated with informant accuracy. More central actors may observe and receive more information about relationships in the group. There has been relatively less work that investigates for which dyads informants can most accurately report tie information (dyad level accuracy). This may be due to limitations of small group size due to most informant accuracy studies using cognitive social structure designs. In addition to overall accuracy, there is evidence that informants have structural biases in self-reports such as a tendency for informants to report themselves as more central than they really are (ego bias) and in third-party reports such as a tendency for informants to overreport transitivity as balance theory would predict (Kumbasar, Rommey, and Batchelder 1994; Krackhardt and Kilduff 1999; Kilduff et al. 2008; Ouellette 2008; Pittinsky 2008; Neal and Cappella 2014; Neal, Neal, and Cappella 2016).

Krackhardt (1990) does not find that betweenness centrality significantly correlates with cogntitive accuracy in advice or friendship networks. He theorizes that this might be because the network was small (36 employees). He also found that closeness centrality was an even weaker predictor of cognitive accuracy than betweenness centrality.

Krackhardt (1987) finds that an individual's self-report indegree in an advice network is negatively correlated with agreement with the consensus network. He calls the strongly symmetrized self-report network the locally aggregated structure (LAS). He also finds that indegree and betweenness in the consensus network are positively correlated with agreement with the LAS.

Johnson and Orbach (2002) find a moderately strong correlation between knowledge of the network and reported centrality in a political network. In the self-report survey respondents rated tie strength from 0 to 10. They call this the reported network R. They use a cognitive social structure survey with a fixed choice design where respondents were asked to list each alter's top three. They call this the individual cognitive matrix X_i for actor i. They measured accuracy as the correlation between reported network R and individual cognitive matrix X_i. They found positive relationship between indegree centrality and accuracy, a tendency for actors to overestimate indegree centrality, and that legislators were more likely to overestimate their centrality than private actors.

Casciaro (1998) finds a positive correlation between degree centrality and accuracy in the perception of both friendship and advice networks. Bondonio (1998) reanalyzes data from Krackhardt (1990) and finds that the only significant predictor of individual level accuracy is indegree centrality. Bondonio distinguishes between invidual level competency and dyadic level competency. At the dyadic level, Bondonio finds that geodesic proximity in the advice LAS and similarity in age (for friendship LAS) and tenure (for advice LAS) positively correlate with accuracy. Simpson and Borch also find that accuracy decreases with geodesic distance (coummunicated to Ouellette and reported in Ouellette 2008). Ouellette 2008 in a meta-analysis of five friendship network cognitive social structures did not find that geodesic distance nor betweenness centrality were significantly associated informant accuracy. However, these networks were quite small, non-response was high, and non-respondents were deleted. Noah E Friedkin 1983 finds that informants are only able to accurately perceive the role performance of others up to a geodesic distance of two and calls this the horizon of observability. The horizon of observability has also been called the perception radius Sewell 2018 or the perceivable proximity threshold Oguz 2014. Kossinets and Watts 2006 finds in an email network of university students that in the absence of a shared class that the probability of tie formation decreases rapidly with geodesic distance. Beyond centrality and social distance there are fewer studies that examine the relationship between network position and informant accuracy. One such example is Daniel, Silva, and Santos 2018 which found that informants "with a higher proportion of transitive relationships are better at perceiving who affiliates with whom, and that increases in transitivity associate with increases in perception accuracy".

Respondents have generally been found to be more reliable in self-reports than in third-party reports (adams and Moody 2007) although this finding is not universal (Neal, Neal, and Cappella 2016). McEvily (2014) found 72-87% accuracy while Krackhardt found 20-40% accuracy (Krackhardt 1996). McEvily states "Although I cannot be certain, I would speculate that the egos in my study were more familiar with and knew better their alters than Krackhardt's 21 managers' familiarity and closeness with each other. If so, a second implication of the difference between our two studies is that the degree of accuracy may vary with the nature and quality of egos' relationships with their alters."

2.1.3.3 Accuracy and Individual Characteristics

Researchers have probed how individual characteristics such as gender and personality relate to informant accuracy in reporting social ties. Casciaro (1998) defines global competency as Phi correlation of informant's cognitive social structure and the locally aggregated structure. She finds that need for achievement and degree centrality have significant positive correlation with global competence, while hierarchical level has a negative correlation. She also finds a positive correlation between need for achievement and accuracy in the perception of both friendship and advice networks. Need for affiliation positively correlates with accuracy in the perception of the friendship network. Casciaro, Carley, and Krackhardt (1999) find that positive affectivity improved the accuracy of the perception of other's friendship ties but hindered the accuracy of perception of one's own personal advice ties.

While not specifically looking at informant accuracy for social network data, Meijs and colleagues find that female gender and popularity are both associated with higher social intelligence (2010). Parker et al. find that GPA positively correlates with interpersonal abilities (Parker et al. 2004). Individuals with high social intelligence and interpersonal abilities likely retain more information about social ties. In addition to lack of knowledge and errors in recall, informant errors can also result from untruthful or careless responses. Cornell, Lovegrove, and Baly find that boys and students with lower test scores are more likely to have invalid responses (Cornell, Lovegrove, and Baly 2014). Cappella, Neal, and Sahu 2012 did not find a significant effect of gender or positive school behavior, but did find significant positive effects for degree centrality and emotional support and a negative effect for class size. Higher accuracy in smaller classrooms is likely related to social distance as discussed in the prevous section.

Most studies of informant accuracy investigate accuracy in perceiving (observer accuracy) but fewer look at accuracy in perceiving (target accuracy). The social relations model (Back and Kenny 2010) is one approach to separating target effects from observer effects. Neal, Neal, and Cappella 2016 studied target and observer accuracy using CSS data collected from 33 second through fourth grade classrooms. They found that pairs of children were more accurately observed when they "involved same-sex, high-popularity, and similar-popularity children"' and that "relationships between pairs of girls were more accurately observed than relationships between pairs of boys". They found higher accuracy in higher grades and smaller classrooms. In terms of observer accuracy they found that girls were more accurate observers than boys.

2.1.4 Summary and Hypotheses

Given the findings reviewed in the introduction, I posit that informant competence relates to both network position and individual characteristics. I expect that informant competence would positively correlate with knowledge of the alters and their relationship, but also the carefulness with which the informant completes the survey.

Hypotheses 2.1 Network position. I posit two hypotheses regarding how network position relates to informant competence.

Hypothesis 2.1.1 Social distance. Respondents will be more competent informants regarding relationships of actors that they are close to in social space. I expect that informants will not possess sufficient knowledge to accurately report on the relationships of individuals that are distant from the informant in social space.

Hypothesis 2.1.2 Centrality. More central actors will be more competent informants. I expect that informants with higher centrality in the network will have overall greater knowledge of third party relationships in the population due to their interactions with other students, through both observation and discussion. I also expect that popular students will have higher social intelligence and will therefore report relationships more accurately.

Hypotheses 2.2 Individual characteristics. I propose two hypotheses relating individual characteristics to informant competence.

Hypothesis 2.2.1 Gender. Girls will be more competent informants than boys. I expect that boys will be less knowledgeable of peer relationships than girls and will also be more careless in completing the survey.

Hypothesis 2.2.2 Academic achievement. High performing students will be more competent informants. I expect high performing students will complete the survey more carefully and will also pay more attention to relationships within the small learning community. High achieving students are also expected to be more socially intelligent and therefore more knowledgeable of peer relationships.

How the concepts are operationalized will be described in the measures section and how the hypotheses will be tested will be described in the methods section.

2.2 Data and Measures

The data from Project 1 described in Section 1.3 is used for this analysis. Here I focus on the Wave 2 survey of students in Cohort 1 of Small Learning Community A (SLC-A). Of the 93 grade 11 students in SLC-A that year, 84 completed the survey. One was removed for invalid responses. These students were surveyed with an ego network census. They were asked self-report name generator questions where they could nominate up to seven alters. I use three of the four self-report questions: speaking about school work, speaking about personal concerns, and hanging out. For each self-report question and alter, the student was asked how often they had that interaction: daily, weekly, monthly, or rarely. Then the student was asked how often each pair of the alters in their ego network spoke to one another: often, sometimes, rarely, or don't know.

2.2.1 Network Definition

2.2.1.1 Network Boundary

The population for this study is considered to be only the SLC-A Cohort 1 students when they were in 11th grade (Wave 2). The student could nominate people outside this population in the self-report survey, but that data is only used in this study to calculate the percentage of outside nominations.

2.2.1.2 Self-report Network Definition

In this study, I use multiple defitions of the self-report network for different purposes. If a student A lists student B in one or more of the positive relation questions in the self-report (school work, personal concerns, hang out) then I consider there to be a directed tie from A to B. I use this directed network to calculate indegree (popularity), outdegree (sociality), and betweenness centrality. From this directed network, I create two undirected networks: one with weak symmetrization (either student nominated the other) and one with strong symmetrization (both students nominated one another). The strongly symmetrized network is also called the locally aggregated structure or LAS (Krackhardt 1990). There is self-report information on four relations, which I wish to summarize to a single network. I use the weakly symmterized network to find latent social position using the KliqueFinder algorithm which I use to calculate social proximity. I use the LAS to calculate degree centrality. The SLC-A Cohort 1 in Wave 2 is graphed in Figure 5.

2.2.2 Measures

2.2.2.1 Social Distance

In order to test Hypothesis 2.1.1 that informants closer in social space to the dyads that they are evaluating will possess more knowledge of those relationships, I must first choose how to measure social distance. One method to place the actors in latent social space is with the KliqueFinder algorithm which clusters the actors into cohesive subgroups and then applies multidimensional scaling (MDS) within and across subgroups to create latent positions.

KliqueFinder assigns actors to subgroups by using a hill-climbing algorithm to maximize in a p_1 network model the estimated increase in the probability of interaction associated with same subgroup membership (Frank 1995). After partitioning the actors into subgroups, they are placed in a 2 dimensional latent social space by applying multidimensional scaling within and between subgroups and then rotating subgroups to minimize the number of ties that cut across the space of a subgroup (Frank 1996). I run KliqueFinder on an unweighted weakly symmetrized network with non-respondents removed. A plot of the results is shown in Figure 7. There are sixteen subgroups. The number of students in each subgroup ranges from 2 to 7 and the median subgroup size is 5.

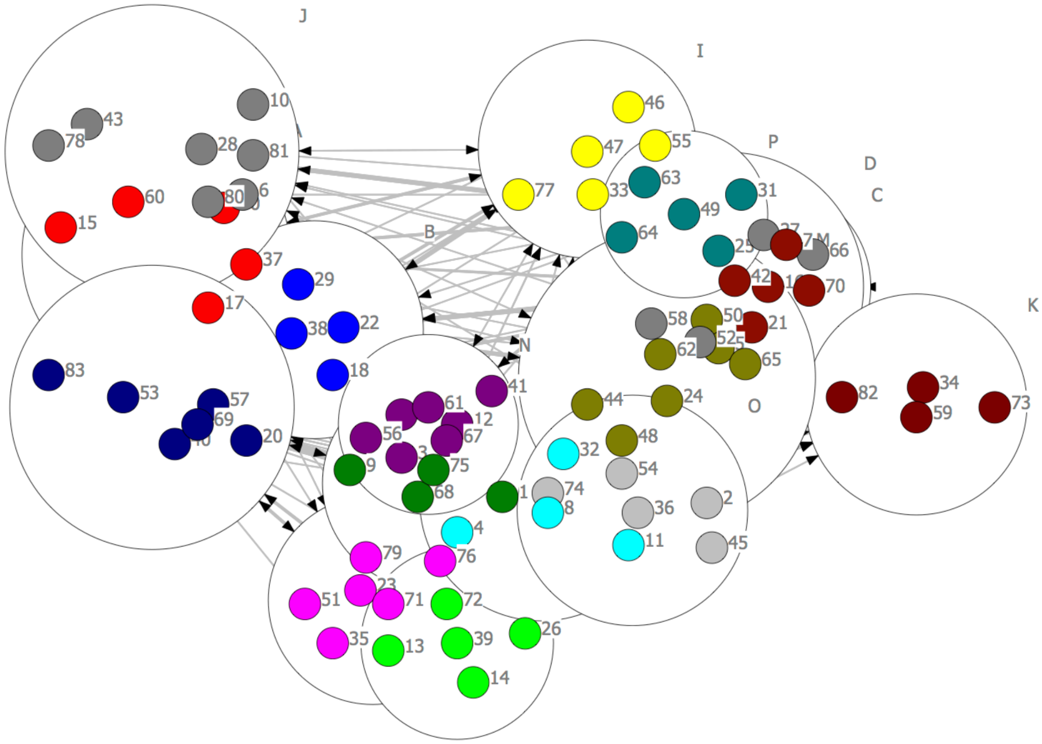


Figure 7. Plot of the KliqueFinder clustering and placement of actors in social space.

To create the measure for social proximity I take the inverse of the average Euclidian distance from the evaluator's latent position to the alters' positions. According to Hypothesis 2.1.1 it's expected that as the proximity to the dyad increases, the evaluator's knowledge of their relationship will also increase.

2.2.2.2 Centrality

According to Hypothesis 2.1.2, more central actors will be more competent informants. There are many way to operationalize centrality. Noah E. Friedkin 1991 explains how centrality measures can be derived from social process foundations. He argues that measures can be complementary and address different questions about social structure, for example by distinguishing between substantive and structural contributions of influence. I will use two measures of centrality, degree and betweenness. In the typology of Borgatti and Everett 2006 degree centrality is a radial measure and betweenness is a medial measure.

The degree centrality of a node is a measure of vertex degree. In an undirected unweighted network, it is simply the number of ties a node possesses (Freeman 1978). Degree centrality is a measure of popularity and sociality. I measure degree in the locally agreggated structure, the strongly symmetrized network. These individuals are more socially active and can be expected to observe more interactions. They are also likely to be more socially intelligent and may retain more knowledge of peer relationships.

The betweenness centrality $C_B(v)$ of node v is defined as

$$C_B(v) = \sum_{i,j:i \neq v \neq j} \frac{g_{ivj}}{g_{ij}}$$

where g_{ivj} is the number of geodesics (shortest paths) from i to j through v, and g_{ij} is the total number of geodesics from i to j (Brandes 2008). Betweenness centrality is related to the concept of brokerage. Informants with higher betweeness centrality broker information between different subgroups and thus it would be expected that they have more received knowledge of third party relationships. I measure betweenness in the weakly symmetrized network. Degree and betweenness are expected to be strongly correlated, so I will test them in separate models.

2.2.2.3 Informant Characteristics

Gender is one of the informant characteristics expected to correlate with informant competence. Unfortunately the survey did not include direct measures of social intelligence. To measure academic achievement, I use the sum of reading and math standardized test scores from the Iowa Test of Basic Skills the students took the second semester of their sophomore year.

2.3 Analysis

The goal of the analysis is to help understand which respondents can best report on which relationships. One of two methods are typically used to evaluate informants in the literature, accuracy or competence. The accuracy of third party reports is evaluated against some data which is treated as the "ground truth." Observation, archival data, or self-reports are all sometimes used as the ground truth. Informant competence is typically measured by the correlation of informants' evaluations with those of the rest of the evaluators. In my data, there was no data that could reliably be used as the "ground truth": there was no observation or archival information on contact frequency in the SLC data and the self-report survey questions do not match the third-party question. Therefore accuracy is not available, so I will use a measure of competence modified to account for the fact that have ego network census instead of CSS data. Unlike a CSS design, in the ego network census there may be few or no other evaluations of a given dyad to compare. To deal with the missing third party reports needed for the typical calculation of informant competence, I propose to instead model the expected third party report given the self-report data. In this way, a submodel of informant precision about the expectation becomes a measure of competency. This is similar to how Sewell 2018 models bias and variance of informant reports. I use a mixed ordinal location scale regression with a random informant effect on location (Agresti 2012; Hedeker, Demirtas, and Mermelstein 2009). The location submodel captures how an informant would be expected to report on a relationship given the self-report data and a scale submodel captures the dispersion of the reports from the expectation given the informant's network location and individual characteristics. Lower dispersion means higher competence. The dispersion of informant reports is expected to be lower if the informant is knowledgeable about the relationship and higher if the informant responds carelessly.

The location submodel includes measures of social proximity of the alters to one another and from the evaluator to the alters, a random effect on the evaluator, and a control for the percent of the evaluator's self-report nominations that are not another student in the small learning community. KliqueFinder is used to find alter-alter and average ego-alter proximities. As an additional measure of alter-alter proximity I use a sum of the six contact frequencies from the three name generators in the self-report.

The scale submodel includes measure of network location and individual attributes. Network measures included are the KliqueFinder average proximity from evaluator to alters and for centrality either degree or betweenness. Individual characteristics included are the student's gender and the sum of reading and math test scores.

The model may be motivated by considering the evaluator to have a latent belief about how often the two alters in the dyad interact with one another. Call this unobserved variable y^*_{ijk} and assume the linear model

$$y^*_{ijk} = \alpha^* + X^T_{ijk}\beta^* + U^*_i + \varepsilon_{ijk}$$

$$\varepsilon_{ijk} \sim \mathcal{N}(0, \sigma^{*2}_{ijk})$$

$$y_{ijk} = \begin{cases} 0 \text{ if } y^*_{ijk} \leq \theta^*_1 \\ 1 \text{ if } \theta^*_1 < y^*_{ijk} \leq \theta^*_2 \\ 2 \text{ if } \theta^*_2 < y^*_{ijk} \end{cases}$$

$$\sigma^*_{ijkl} = e^{z^T_{ijk}\zeta}$$

U^*_i is the latent random informant effect.

Then the cumulative probabilities of the observed variables are

$$y_{ijk} = \mathrm{P}\left(Z \leq \frac{\theta^*_l - \alpha^* - X^T_{ijk}\beta^* - U^*_i}{\sigma^*_{ijk}}\right) = \Phi\left(\frac{\theta_l - X^T_{ijk}\beta - U_i}{\sigma_{ijk}}\right)$$

with relative scale

$$\sigma_{ijk} = \frac{\sigma^*_{ijk}}{\sigma^*} = e^{z^T_{ijk}\zeta}.$$

The latent variable y^*_{ijk} is assumed to be normally distributed so the link function is probit. Latent scale parameters are not identifiable, so parameters are estimated for the observed variables. The model is fit using maximum likelihood estimation.

2.4 Results

Table 7 presents the regression results for Models 2.1 and 2.2. The inputs have been standardized by mean centering and dividing by 2 standard deviations so that the coefficients are more easily compared (Gelman 2008).

I find that as the evaluator gets closer to the alters in the dyad evaluated, dispersion decreases, as predicted by Hypothesis 2.1.1. A 2 standard deviation increase in average KliqueFinder proximity from the evaluator to the alters is associated with about a 25% decrease in dispersion. For comparison, I fit models using alternative measures of social proximity. In Models 2.1.2 and 2.2.2, I replace the KliqueFinder proximity with a social proximity calculated by applying nonmetric multidimensional scaling directly to the geodesic network distance. In Models 2.1.3 and 2.2.3, I measure social proximity with a binary indicator for being in the same KliqueFinder subgroup or not. Results for these models are shown in Appendix C.2.2. KliqueFinder proximity which is derived by applying MDS within and between subgroups could be viewed as intermediate between the binary measure of same KliqueFinder subgroup and the fully continuous measure derived by applying MDS directly to geodesic distance. The coefficient estimates on social proximity in the scale submodel have slightly larger magnitude with MDS proximity than KliqueFinder proximity and have the smallest magnitude for same subgroup proximity. Models 2.1 and 2.2 using KliqueFinder proximity have higher likelihood than Models 2.1.2 and 2.2.2 with MDS proximity. In short, although the coefficient magnitudes and overall model fits vary slightly depending the exact measure of social distance utilized, the results remain very consistent.

The data do indicate that more central actors are more reliable informants as Hypothesis 2.1.2 posits. I find that higher scores in degree centrality and betweenness centrality both correspond to lower dispersion in third party evaluations. A 2 standard deviation increase in LAS degree is associated with about a 25% decrease in dispersion. A 2 standard deviation increase in betweenness is associated with about a 22% decrease in dispersion.

The results suggest that individual characteristics may significantly predict informant competence. I find evidence for Hypothesis 2.2.1 that girls are better informants. Boys have about 22% higher dispersion in third party evaluations than girls, all else being equal. Hypothesis 2.2.2 that high academic achievers are more competent informers on relationships within the small school. A 2 standard deviation increase in test scores is associated with about a 27% decrease in dispersion.

Table 7. Cumulative link models of third party evaluations with standardized coefficients.

	Model 2.1		Model 2.2	
	Estimate	p-value	Estimate	p-value
Scale Submodel				
Mean KF proximity to alters	-0.320	0.003	-0.263	0.017
Betweenness	-0.251	0.005		
LAS degree			-0.286	0.001
Male	0.208	0.049	0.198	0.055
Test scores	-0.327	0.001	-0.301	0.002
Location Submodel				
Alters nominations sum	0.292	0.000	0.272	0.000
Alters KF proximity	0.501	0.000	0.491	0.000
Mean KF proximity to alters	0.045	0.730	0.053	0.674
Percent outside noms	-0.454	0.030	-0.433	0.036

2.5 Discussion

This project uses a novel data set and modeling approach to identify correlates of informant competence based on individual characteristics and network position. The question of informant accuracy is central to the study of social networks and there is a need for further research on the topic, particularly in the context of moderate to large networks.

The results of this study provide additional evidence regarding previously investigated relationships between informant competence and network centrality. There is fairly strong evidence of a positive correlation between informant competence and degree centrality (Casciaro 1998; Bondonio 1998; Johnson and Orbach 2002; Cappella, Neal, and Sahu 2012). Evidence is somewhat weaker that betweenness centrality is associated with higher informant competence (Krackhardt 1987, 1990; Ouellette 2008).

Here I have found that in a contact network of high school students both degree and betweenness centrality are positively associated with informant competence.

Previous studies that have looked at the relationship between social distance and informant competence have typically used geodesic, or shortest path, distance (Bondonio 1998). I have used a different method of calculating social distance. I cluster individuals into cohesive subgroups using the KliqueFinder algorithm and assign them to latent social positions by mapping the interactions within and between subgroups via multidimensional scaling (Frank 1996). I then calculate social distance based on these latent positions. Given the importance of cohesive subgroups in structuring interactions, this operationalization could represent a more valid measure than geodesic distance for capturing the extent to which an individual has opportunities to learn about an alter. If so, with regards to understanding informant accuracy, it then offers a more powerful tool to identify the horizon of observability. That a horizon of observability for informant accuracy in reporting social ties exists is of course not in dispute, but previous work has not been able to effectively explore the measurement of social distance and its relationship with informant accuracy due to the reliance on full CSS designs and the consequent limitation in the size of the networks for which data has been collected. Equipped with a new measure of social distance and data from a larger population in comparison to previous studies, I have demonstrated that informant competence does significantly decrease with distance.

The relationship between informant competence and individual characteristics is less well established. In particular studies regarding gender and informant competence have had somewhat mixed results (Neal, Neal, and Cappella 2016). However, the overall tendency has been towards finding that girls are better informants and my results offer additional support for that hypothesis. There is even less evidence regarding the relationship between academic achievement and informant competence. Malloy, Albright, and Scarpati 2007 found that academic achievement is positively related to competence in perception of individual attributes of peers but did not measure perceptions of social ties. Cappella, Neal, and Sahu 2012 hypothesized that students with positive school behavior would have higher informant competence in perception of social structure, but did not find a significant effect after controlling for other factors. Thus the finding in this study that students with higher test scores were more competent informants offers a valuable addition to the body of work on informant accuracy. High performing students might have better knowledge of the social structure as a result of greater social capital, social-cognitive ability, or other factors. They might more accurately report their knowledge for reasons such as greater task persistence. Further inquiry may help to elucidate the underlying causal mechanisms of the relationship between academic achievement and informant accuracy.

As informant accuracy may depend on individual characteristics, network position, and network topology more investigation is needed to explore how robust currently employed methods are to such dependencies.Butts 2003 conducts a Monte Carlo test which finds that although the Butts' Bayesian Network Accuracy Model does not implement structurally correlated accuracy parameters it appears to still be quite robust to centrality-correlated accuracy. However, I note two reasons why robustness might be expected to be lower in other contexts. First, the test correlated informant accuracy parameters with the informant's centrality in the true network, but the model was also estimating accuracy parameters for each informant. Such a model should be able to easily handle any structurally correlated errors which have a constant effect for all of a given informant's reports. The model might not be so robust in cases where informant accuracy has dependencies other than on actor covariates of the informant.

If the errors were instead correlated with social distance then the effect would differ depending on the dyad that the informant is evaluating. If informant reports show correlation of error in cognitive structure, for example through a stronger tendency to perceive balance at close and far social distances as Krackhardt and Kilduff 1999 finds, then the effect would also not be constant per informant. Second, the test used simulated cognitive social structure data with 15 actors and thus 15 observations per dyad. If an arc sampling design were used with fewer observations per dyad the results might also be less robust.

It is worth noting that an effect of social distance on informant competence was found in spite of the limitation of third-party reports only coming from the personal networks of students in a small learning community. Due to the use of the ego network census design rather than a cognitive social structure or balanced k-arc sampling, informants were only asked about dyads where they had reported having a tie to both individuals. This limits the social distance from the evaluator to the alters. Higher error rates and proportion of don't know responses would be expected outside of the ego network. A similar study with a balanced k-arc sampling design could be used to explore the limits of when informants are able to accurately report on dyads outside their ego network.

In interpreting my findings, two broad limitations of the model estimation and data warrant mention. In terms of the model estimation, my substantive question of interest required explicitly modeling the variance of the dependent variable. Approaches to deal with such "heteroscedasticity" when it is simply considered a nuisance to causal inference are well-developed. However, models that treat it as an object of interest are more novel, and it should be noted that their properties, even when using very large samples, are not yet well-understood (Keele and Park 2006).

With respect to the data, as I reviewed in the introduction, BKS and many others have shown that self-reports are not completely accurate. Additionally, the third party reports are not missing at random. There is a potential sampling bias in that survey respondents provided evaluations only for the dyads in their self-report ego network (Butts 2008a). Explicitly modeling such error and non-missing randomness was beyond the scope of this analysis, but one could imagine an approach that might take this into consideration (An and Schramski 2015). For example, investigators have demonstrated various systematic biases in social perception, such as informants reporting themselves as more central than they are (Kumbasar, Rommey, and Batchelder 1994; Johnson and Orbach 2002; Neal and Cappella 2014) and informants perceiving the network as more clustered than it actually is (Krackhardt and Kilduff 1999; Kilduff et al. 2008; Ouellette 2008; Pittinsky 2008; Neal, Neal, and Cappella 2016). In this analysis, the random informant effect in the location submodel can capture individual biases in their average response which corresponds to density but this does not account for other potential structural biases.

In addition to work addressing these limitiations, there are several additional open questions related to network measurement which could be pursued. First, survey respondents sometimes respond carelessly, especially if the survey becomes tedious. When completing an ego-network survey, the number of third-party dyads that a respondent is asked to evaluate grows quadratically with the number of alters nominated in the self-report. When this number grows large, respondents are more likely to provide spurious answers (McCarty, Killworth, and Rennell 2007). In a computer survey they may try to click through as quickly as possible without responding attentively (Matzat and Snijders 2010). Is there evidence of this behavior? If so, how can this source of error be modeled to improve the accuracy of network measurement?

Second, name generators are free recall questions. What can we learn from the order and timing of the nominations? Respondents may recall alters in groups (Freeman, Romney, and Freeman 1987). For example, a respondent might first list the friends that they know from middle school, then friends from marching band, followed by friends from German class, and so on. If that were the case, then we would expect that, on average, actors that are closer together in social space would be nominated closer together as well. Respondents might recall close friends before acquaintances. Such observations might be used to provide more accurate estimates of ties or tie strength.

Third, the ego net approach might leave many dyads without any third party evaluations. However, the K-replication balanced arc sampling method may have problems with lack of knowledge of distant dyads. Could we develop a new measurement methodology that provides better coverage while asking respondents to evaluate dyads that they are likely to know? Besides individual characteristics available from roster data, are there other ways to more accurately identify good informants? Could we draw on research on social cognition and social intelligence? Could we better take advantage of possibilities of computer surveys to obtain more accurate measurements, perhaps by prioritizing the questions asked based on updated calculations of uncertainty using earlier answers by the respondents?

Informant accuracy is likely to continue to be a fruitful area of research in the field of social networks. It is my hope that this will contribute to the development of improved methods for network measurement that have greater validity and efficiency thereby providing better tools to advance understanding of human sociality. In the following project, I look at how information from third-party reports can improve predictive accuracy in missing tie imputation.

Chapter 3

PROJECT 3: IMPUTATION FOR MISSING LINKS

When collecting network data, it is desirable to have a network census – i.e., as close to complete response as possible. Due to the relational nature of the data itself and approaches to modeling it, missing data is more problematic than for studies that focus on the individual. Unfortunately, avoiding non-response entirely may be unrealistic. Both cost and access are often insurmountable issues. The most common methods for dealing with missing data in networks are deletion-based. Studies suggest that model-based imputation methods perform better. However, such work is still in the early stages of development and much remains to be know about how to best incorporate model-based imputation in the analysis of network data.

In this paper, I exploit the use of a promising source of information for missing network data imputation: the reports of connections between two individuals by third parties. More specifically, I examine the extent augmenting current ERGM-based imputation methods with third-party information can improve imputation accuracy. My findings give cause for optimism: models incorporating third-party reports consistently outperformed traditional specifications in accuracy, especially when non-response is high.

In Section 3.1, I review the existing literature on general approaches to dealing with missing network data, as well as the specific are of imputation of network data. In Section 3.2 and 3.3, I describe my data and analytic strategy. Section 3.4 presents the results. Section 3.5 concludes with implications for current research and future network data collection.

3.1 Background

3.1.1 Sources of Network Measurement Error

Missing data may be categorized within a broader scheme of research design errors (Znidarsic, Ferligoj, and Doreian 2012). In particular, missing edge data from actor non-response is closely related to other sources of error such as boundary specification problems, fixed choice designs, and informant inaccuracy (Kossinets 2006; Feld and Carter 2002; Bell, Belli-McQueen, and Haider 2007). When analyses are restricted to complete cases (respondents only) then the problem of non-response is equivalent to a boundary specification problem with the non-respondents excluded. With a fixed choice design, when a respondent nominates the maximum number of alters then censoring creates uncertainty about ties to the remaining alters as it is unknown if the respondent would have nominated them given the opportunity. The typical treatment of fixed choice designs is to impute the unconditional mean by assuming that the uncertain ties are absent because the network is sparse. This corresponds to the commonly used null-tie imputation for non-respondents where non-respondents are treated if they had no outgoing ties. Similarly, free recall in name generator questions leads to errors created by forgetting of alters and this also introduces false negatives. This helps to explain why false negative rates are generally much higher than false positive rates in network data collected via free recall. Attempts to estimate the true network (or the criterion graph in Butts' terminology) in the presence of missing data, other sources of error, or both are all efforts at network reconstruction. Studies that focus on the identification of false positives and false negatives within an observed network such as Clauset, Moore, and Newman 2008 and Guimera and Sales-Pardo

2009 are therefore still quite relevant to the development and evaluation of methods for missing data imputation. These tasks, also called denoising and completion, may be combined (Peixoto 2018). In this study I focus on missing data from actor non-response.

3.1.2 Treatments for Missing Edge Data

Missing data in network studies can be handled by

1. complete case analysis,
2. using an inference method that appropriately handles missing data, or
3. imputing the missing data and performing inference with the imputed data (Huisman and Krause 2017).

The most commonly used method is complete case analysis where actors with missing data are removed from the analysis. This is also called case deletion, listwise deletion, or respondents only analysis and it is the treatment that was used in Project 1. The second approach is called the model-based approach. Exponential random graph models can appropriately handle missing data provided that it is ignorable (Koskinen et al. 2013). In addition to providing inferences for model parameters without separately imputing the missing data first, likelihood-based or Bayesian procedures can be used for imputation as well by drawing the missing data from its distribution conditional on the observed data. Model-based imputation can be based on models such as ERGMs (C. Wang et al. 2016; Hipp et al. 2015), Kronecker graph models (Kim and Leskovec 2011), feature models (Miller, Jordan, and Griffiths 2009), latent position models (Hoff 2009), and stochastic block models (Guimera and Sales-Pardo 2009; Peixoto 2018).

Simpler methods of imputation include

1. treating all outgoing ties from non-respondents as absent,
2. assuming that non-respondents reciprocate all incoming ties, or
3. relying on third-party reports to estimate a consensus structure.

Also called null-tie imputation, the first method imputes the unconditional mean as there are more absent edges (null ties) than present edges. Null-tie imputation is the most common approach after complete case analysis. Imputation by assuming reciprocity is called reconstruction by Stork and Richards (1992a). These two methods clearly introduce a bias in the imputed data. Reconstruction by reciprocity also does not help when both members of the dyad are non-responders.

Reconstruction by reciprocity combines observed data with an assumption about the structure of the network. It would be similar to inference of the missing data in an exponential random graph model where the only covariate is reciprocity and the prior is sufficiently strong to overwhelm the data. An ERGM which does not use such a strong prior could use the observed data to estimate the actual level of reciprocity rather than assuming that incoming ties are either never or always reciprocated. Imputation performance would most likely be further improved by including other effects such as transitivity as was found by Wang et al. (2016).

Third-party information can be collected via a cognitive social structure (CSS) design where informants are asked to report on the all dyads in the population. Such CSS data can be aggregated by simple rules such as a majority rule to impute missing data from non-respondents (Neal 2008). Third-party reports can also be collected in a method called social cognitive mapping where informants are not asked about every dyad but are instead asked to assign alters to subgroups (Cairns et al. 1989). The co-occurrence of individuals in reported subgroups can be used as a method of imputing tie data for non-respondents. Model-based procedures that incorporate third-party reports are possible as well. In 2003 Butts proposed the Butts' Bayesian Network Accuracy Model to improve network measurement by drawing on information from multiple observations per dyad to estimate false positive and false negative error rates. The multiple observations could just be two self-reports from the members of the dyad but they could also include third-party reports. Butts has made available an implementation of this model in the R package sna with its bbnam function (Butts 2008b). I have found only two published usages of the method in the literature: Marcum, Bevc, and Butts 2012 in which the pooled error model was used with a Bernoulli graph prior to analyze a self-report network with a relatively high error rate and Lee and Butts 2018 using the per observer error model with a continuous mixture of $U \mid man$ graphs as the network prior to analyze cognitive social structures (Butts 2017). Butts suggests collecting third-party reports using a K-replication balanced arc sampling design when the population is large enough that a cognitive social structure would not be feasible. So far this advice has yet to be embraced by the social network research community. As Peixoto 2018 finds after reviewing network catalogs including the Koblenz Network Collection (KONECT) (Kunegis 2013) and the Colorado Index of Complex Networks (ICON) (Clauset, Tucker, and Sainz 2016) the vast majority of network data is reported as a single adjacency matrix with no error estimates.

The statistical relational learning and link prediction literature also overlap with research on network reconstruction (completion and denoising) (P. Wang et al. 2015; Liben-Nowell and Kleinberg 2007; Lu and Zhou 2011; Brugere, Gallagher, and Berger-Wolf 2018; Kim and Leskovec 2011; Taskar et al. 2007; Jensen, Neville, and Gallagher 2004; Taskar et al. 2003). The term link prediction is most often used to refer to the problem of predicting the formation of new links and in this sense link prediction is a missing data imputation problem where the missing data is a the network at a later time point. However, the term is also sometimes used to refer to network reconstruction a cross-sectional setting. Recently Newman 2018 and Peixoto 2018 have developed methods in the vein of Butts 2003 that also combine a network model with a data model to estimate network structure from noisy data. Peixoto's work has focused on the combination his data model for noisy network measurement with a hierarchical stochastic block model for the network prior and he has made available an implementation in his open-source graph-tool package (2014). This study attempts to contribute to this work by evaluating a model-based imputation that combines self-report and third-party information.

3.1.3 Evaluation of Imputation

Imputation is most commonly assessed by the extent to which the imputation procedure avoids bias and reduces variance for the estimates of particular statistics at the node level, such as centrality, or at the network level, such as transitivity (Krause et al. 2018). In contrast Held-Out Predictive Evaluation (C. Wang et al. 2016) is a cross-validation method that assesses an imputation procedure's accuracy in reproducing the missing edge data itself. A set of observed edges are selected to be held-out, or treated as unobserved, along with the actually unobserved edges. The imputation procedure is then performed on this training data and accuracy is measured against the test data – the edges in the hold-out set. These steps are repeated to produce the cross-validation results. The held-out edges could be sampled directly (tie non-response) or actors could be sampled to have all of their nominations withheld (unit / actor non-response).

The HOPE method provides a more general and stringent benchmark than the usual approach to assessing imputation quality. HOPE is more general because it is not tied to a particular analysis. It is more stringent because while many possible imputations may produce the same parameter estimates for a given model, only one imputation is correct in the sense of recovering the true values and that is what HOPE evaluates. This is not to say that HOPE is better than the usual approach. When it is known what analyses will be performed, it makes sense to evaluate imputation with respect to those same analyses. Recovering the exact missing values may not be important in such cases. However HOPE is certainly useful as a method for the general comparison of imputation procedures and of network model fit. Unfortunately, it has yet to gain significant traction. Outside of the original 2016 paper in which HOPE was introduced, so far there are few works that cite and use the method (Phillips 2017; Zhang and Butts 2017). There are some studies that have used essentially the same predictive evaluation procedure without calling it HOPE. Some were before 2016 (Taskar et al. 2003); others perhaps were unaware of C. Wang et al. 2016 (Peixoto 2018).

Regardless of what it is called, when the method has been used performance has most often been evaluated with tie non-response. Actor non-response is much more prevalent than tie non-response in actual social network survey data and it is also presents a more dificult condition for network reconstruction than tie non-response. The focus of this study is on the general utility of third-party information for missing edge imputation so I will use the Held-Out Predictive Evaluation method to evaluate performance. I will use actor non-response for HOPE since it is more relevant to the context of social network surveys. In C. Wang et al. 2016 the authors footnote that it might be useful to perform HOPE with specific subsets of edge variables. I will disaggregate HOPE results by edge subsets to explore in more detail how third-party information impacts predictive performance.

3.2 Data and Measures

I use the Wave 2 data from the SLC-A Cohort 1 students. This data is described in Projects 1 and 2. This project uses the directed network. Homophily covariates are gender, ZIP code, and fourth semester cumulative GPA. Binary match is used for categorical covariates gender and ZIP code. Similarity as defined in Project 1 is used for continuous covariate GPA. I use as a covariate the average third party evaluation. The number of third party evaluations is the number of ego networks in which the dyad is found. Most dyads do not have any third party evaluation. This is missing data which I impute with the average. This is not a very good imputation method but since the data is not missing at random, I also include a binary covariate for whether there was at least one third party evaluation and therefore the average was not missing.

There is unit non-response due to students missing the survey or being removed for invalid responses. There is no item non-response, so the outgoing edges of each actor are either completely observed or completely missing. Of the 93 students in the SLC, there are nine students that did not complete the survey and one that was removed for invalid responses. This is an overall 11% non-response.

There is evidence that non-response is not at random. This is expected because the primary reason for non-response was being absent the day of the survey.

3.3 Analysis

To evaluate the utility of third-party information for missing edge imputation, I use an exponential random graph model-based multiple imputation procedure and compare the imputation accuracy of a model that includes the average third-party evaluation versus one that does not. I calculate imputation accuracy using the Held-Out Predictive Evaluation (HOPE) cross-validation method proposed in (C. Wang et al. 2016). I compute accuracy for different subsets of the held-out edges in order to better understand the contribution of third-party information.

Exponential random graph models are the most common choice of model for social networks and can capture widely observed and dependent network structures such as homophily, mutuality, and triadic closure. ERMGs have been shown to perform well for network imputation, yielding less bias than commonly used methods (C. Wang et al. 2016; Krause et al. 2018). For these reasons I choose to use ERGMs in this project by comparing a base ERG model that does not include third-party evaluations to an augmented model that does. The probability of an exponential random graph Y may be written as

$$P(Y) = \frac{\exp(\theta^T z(Y))}{c(\theta)} \tag{3.1}$$

where θ is a vector of model parameters, $z(Y)$ is a vector of sufficient statistics, and $c(\theta)$ is the normalizing constant. Following the notation of (C. Wang et al. 2016) when there is missing data the likelihood may be written as

$$P(Y^{obs} = y^{obs}) = \frac{\sum_{y^{mis} \in \mathcal{Y}^{mis}(y^{obs})} \exp(\theta^T z(y^{mis} \cup y^{obs}))}{c(\theta)} \tag{3.2}$$

where y is the state of the random graph Y, y^{obs} is the observed part of y, and $\mathcal{Y}^{mis}(y^{obs})$ is the set of possible imputations of y^{mis}. I perform maximum likelihood inference for

θ using the ergm package in R (Hunter et al. 2008). Using the maximum likelihood estimate of the model parameters I then simulate networks conditional on the observed data to obtain imputed values for the missing edges while keeping the observed edges fixed.

Studies show that ERGM imputation performance depends on the specification being sufficiently complex to fit the data well (C. Wang et al. 2016; Krause et al. 2018). I propose Model 3A as the base ERGM without third-party evaluations. This specification includes effects for gender; homophily on gender, ZIP code, and GPA; triadic closure (GWESP with decay parameter 0.1); and an indicator for whether there was at least one third-party evaluation. Model 3B includes the same effects as Model 3A but adds an effect for the average third-party evaluation. The missing values for the average third-party evaluation are imputed with the overall mean of this dyadic attribute.

In order to evaluate the performance of the above imputation procedure, I use a cross-validation method called Held-Out Predictive Evaluation (C. Wang et al. 2016). As we cannot know whether the actual missing edges are imputed correctly, this method introduces additional misssing data and evaluates the imputation accuracy for these held-out edges. The SLC data only has actor non-response without any tie non-response so this is the type of missingness I introduce with the HOPE procedure. Actor non-response is also the most common type of missing data in social network surveys so the findings will have more external validity using this missingness pattern. I sample holdout sets with an additional 1, 4, 9, 13, 18, and 22 respondents selected uniformly at random to be treated as non-respondents for Held-Out Predictive Evaluation. This corresponds to an overall level of non-response of 12%, 15%, 20%, 25%, 30%, and 34% respectively. The holdout sets are sampled 80, 150, 130, 110, 90, and 70 times respectively. The ERGM models with and without the average third-party evaluation effect are estimated for each holdout set. I impute 250 networks for each model fit using ERGM simulation of the unobserved ties (truly unobserved plus the held-out ties) conditional on the observed ties.

Using the imputed networks, I evaluate imputation accuracy for the six levels of additional non-response on several subsets of edges. I separate the accuracy of reproducing present edges (true positive rate) versus absent edges (true negative rate) as there are many more absent than present edges. I then look at the accuracy only for the subset edges that do have at least one third-party evaluation since this project is concerned with the performance benefit of that information. Lastly, I further focus in on the edges where there is no reciprocal edge information because we might expect the third-party information to be particularly valuable in that case.

3.4 Results

Using the Held-Out Predictive Evaluation approach to assess ERGM-based imputation with and without third-party information, I find that adding an effect for the average third-party evaluation to the model increases the accuracy of imputation for both present and absent edges. These results are presented in Figures 8 and 9.

As in C. Wang et al. 2016 imputation accuracy decreases as the proportion of missing actors goes up, but the accuracy decreases at a lesser rate when using the model that includes third-party information. There are 10 missing actors in the dataset out of 93 total (10.8%) and I add an additional 1 to 22 missing actors for the HOPE procedure. As the percentage of missing actors increases from 11.8% to 34.4% the accuracy of reproducing present edges declines from 42.4% to 35.1% for the model

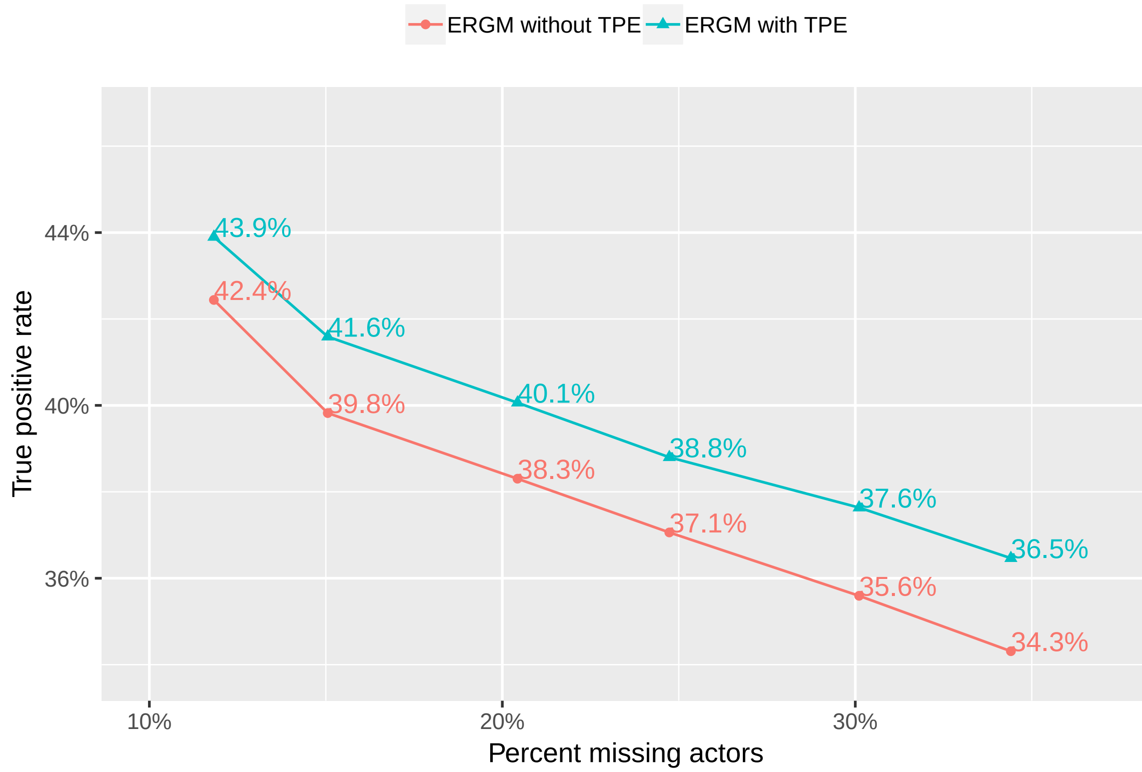


Figure 8. Accuracy of reproducing present edges

without third-party information and from 43.9% to 37.4% when it is added. Thus for present edge imputation the accuracy gained by including third-party information widens from 1.5% to 2.3% as the proportion of missing actors goes up. The accuracy of reproducing absent edges declines from 96.1% to 95.6% for the model without third-party information and from 96.3% to 95.9% for the model with it. The advantage of including third-party information for the accuracy of absent edge imputation increases from 0.2% to 0.3% as the proportion of missing actors goes up.

These increases seem small, but they understate the utility of the third-party information due to the paucity of third-party observations in this dataset. Only 1 out of 5 (21%) of the dyads have at least one third-party observation. This is due to the use of an ego network census sampling design as opposed to a cognitive social structure or

Figure 9. Accuracy of reproducing absent edges

K-replication balanced arc sampling design. To further explore the benefit of incorporating third-party information I evaluate the accuracy of present edge imputation on the subset of edges where there is at least one third-party observation. These results are presented in Figure 10.

The accuracy of present edge imputation with both models is higher on this subset of edges, but higher still for the model with third-party information. The accuracy gained from the inclusion of third-party information goes from 2.5% to 4.1% as the number of missing actors is increased. As expected, the benefit of using a model which incorporates third-party information is higher when imputation is evaluated with only the edges where there are third-party observations.

Reciprocal edge information is quite useful for imputing missing edges given the

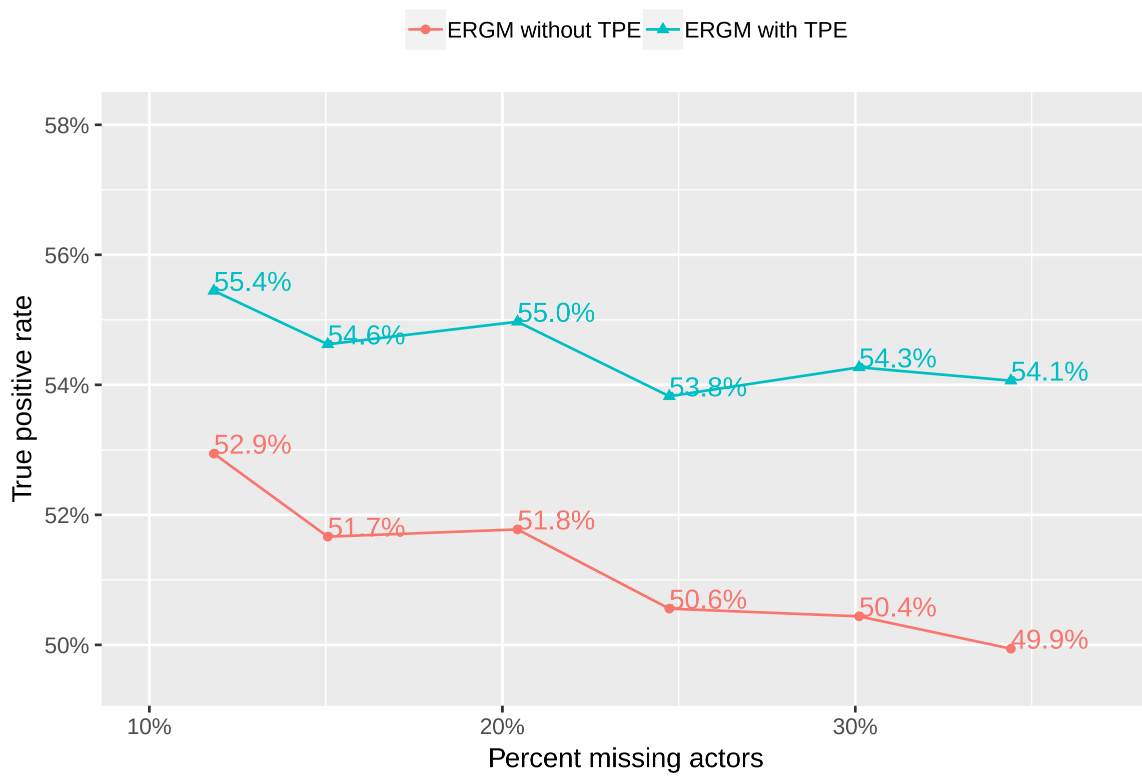


Figure 10. Accuracy of reproducing present edges when there is at least one third party evaluation

high levels of reciprocity present in most social network data. This can be observed in Figure 1 from C. Wang et al. 2016 where a large increase in imputation accuracy is achieved by adding reciprocity to the baseline ERGM. In fact, imputation by reconstruction where missing edges are simply replaced with the incoming edge information has even been suggested as an imputation procedure (Stork and Richards 1992b; Huisman 2009). Third-party information might be especially valuable for edges where the information from the reciprocal edge is not availabe because it is also missing. To test this I evaluate the accuracy of present edge imputation on the subset of edges where not only is there at least one third-party observation but where the other edge in the dyad is also missing. The results are presented in Figure 11.

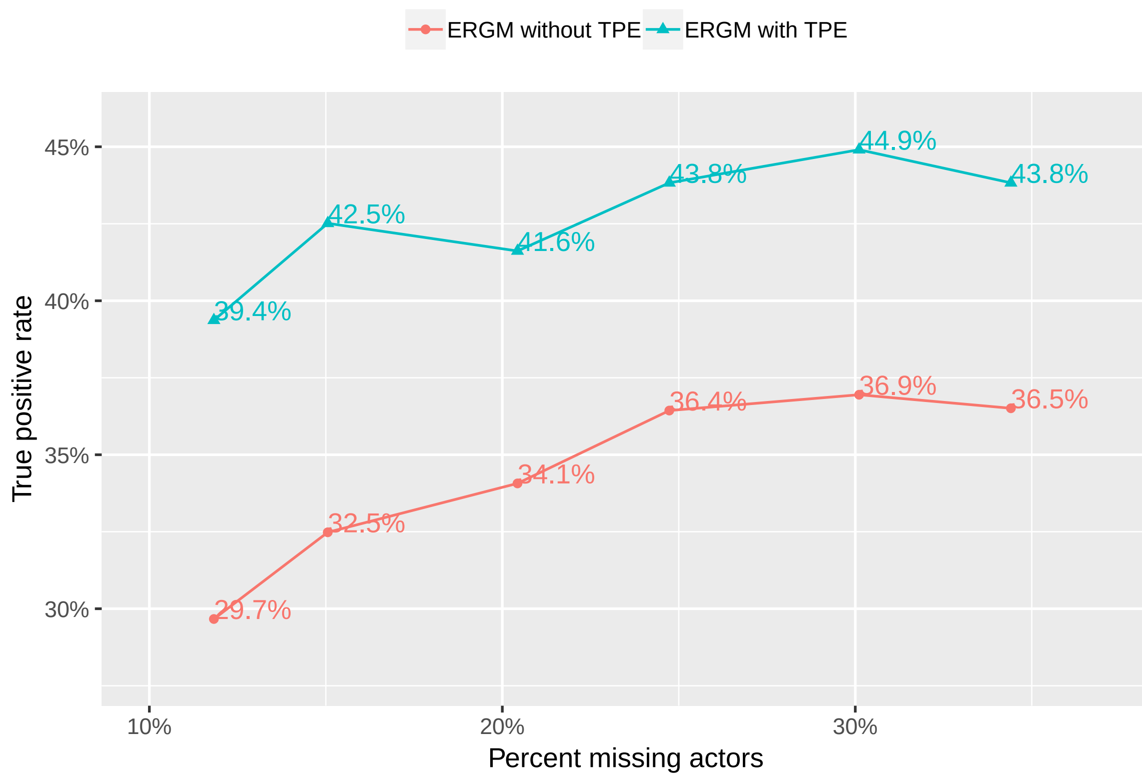


Figure 11. Accuracy of reproducing present edges when there is at least one third party evaluation and the other edge in the dyad is also missing

The increase in accuracy from including third-party information is indeed much higher when further restricting the edge set from those with at least one third-party evaluation to those which are also missing reciprocal edge information. The gain in accuracy ranges from 9.7% to 7.6% as more missing actors are added. Third-party information might be increasingly valuable as the proportion of missing actors increases because so too does the proportion of missing edges for which the other edge in the dyad is also missing. This could explain why imputation accuracy decreases at a lesser rate when using the model that includes third-party information in Figures 8, 9, and 10. In Figure 12 I present the accuracy of present edge imputation on the subset of edges where there is at least one third-party observation and the other edge in the dyad is observed (not missing).

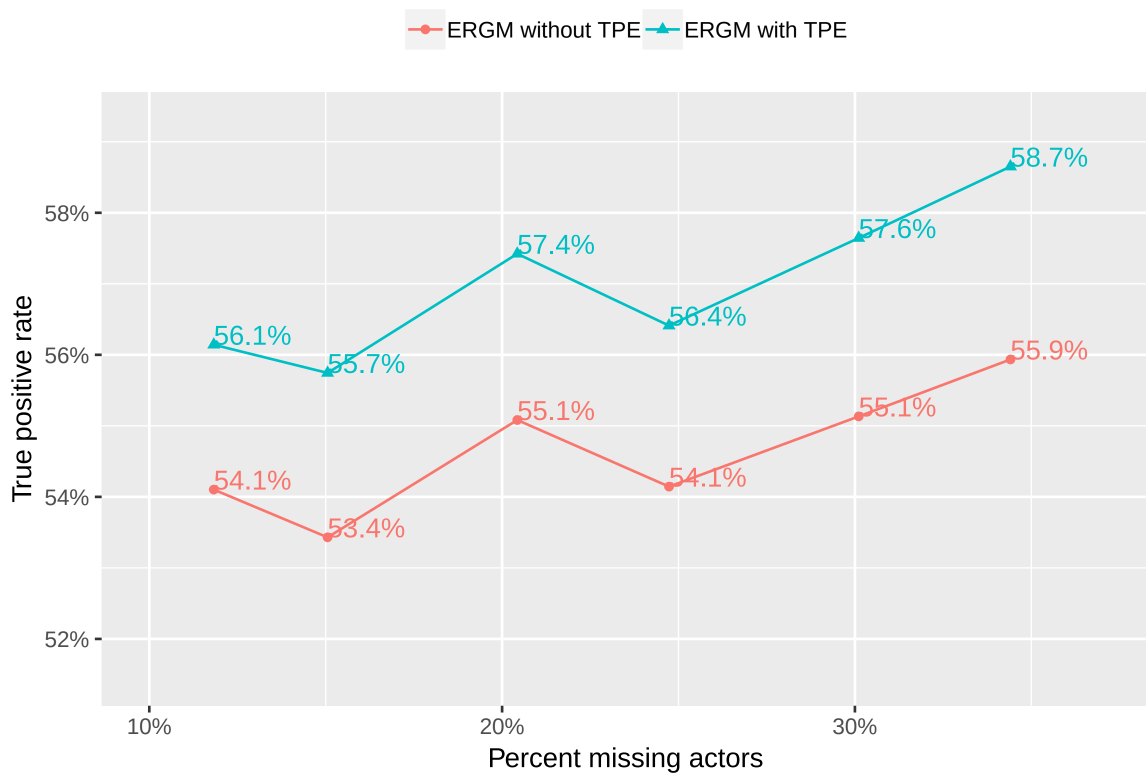


Figure 12. Accuracy of reproducing present edges when there is at least one third party evaluation and the other edge in the dyad is not missing

In both Figures 11 and 12 (where accuracy is evaluated on subsets of the edges with third-party evaluations that do not have and that do have reciprocal edge information, respectively) the gap between the accuracy of present edge imputation for the model with third-party information and the model without decreases as missingness goes up, but this gap increases in Figure 10 (where accuracy is evaluated on the subsets of all edges with third-party evaluations). This is a case of Simpson's paradox. As the proportion of missing actors increases, the proportion of missing edges to be imputed for which the reciprocal edge is also missing increases linearly. The proportions are approximately the same. If NTOTAL is the population size and NMISS is the number of non-responders then the proportion of missing edges for which the reciprocal edge is also missing is

$$\frac{\text{NMISS} \times (\text{NMISS} - 1)}{\text{NMISS} \times (\text{NTOTAL} - 1)} = \frac{\text{NMISS} - 1}{\text{NTOTAL} - 1} \simeq \frac{\text{NMISS}}{\text{NTOTAL}}.$$

These findings support the hypothesis that third-party information becomes increasingly valuable as non-response goes up.

The present results with SLC data may be further compared to the findings in C. Wang et al. 2016 with Add Health data. Both studies perform a predictive evaluation of an ERG model-based imputation procedure with varying levels of actor non-response using data from a social network survey of high school students. The SLC data in this project has 93 students. There is a large range of school sizes in Add Health, so I will focus on the results for the schools with 50 to 150 students for comparison. Out of 93 SLC students there are 83 respondents – the non-response rate is 10.7%. The Add Health schools with 50 to 150 students have a non-response rate that ranges from 3.1% to 22.5% with median 14.8%.

When one holdout is used for the HOPE procedure the true positive rate (accuracy of reproducing present edges) is 42.4% for model 3A and 43.9% for model 3B. These values are at the low end of the range of the true positive rates achieved by the full model in C. Wang et al. 2016 for Add Health schools size 50 to 150 (42.6% - 63.2%, median 50.6%), but are at the high end of the range with the second most complex model (range 22.2% - 47.6%, median 27%). There are two advantages for the performance of the full Add Health model relative to the SLC models in this comparison. The first is that the Add Health dataset and corresponding model are somewhat richer. The Add Health models 4 and 5 include homophily effects for shared classes, clubs, and sports-teams and these covariates are not available in the SLC data. The figures for accuracy with additional missing actors in C. Wang et al. 2016 only display results for model 5 (the full model), so the comparison with models 4 and 5 is taken from figure 1 where the HOPE procedure added random tie non-reponse but zero additional actor non-response.

It can be seen in figure 3 of C. Wang et al. 2016 that the drop in accuracy is much larger going from zero to five percent additional missing actors than subsequent increases of five or ten percent additional missing actors. In fact for most cases the gap from zero to five percent additional missing actors is larger than the gap from five to thirty. There is information available in cases of tie non-response not available for actor non-response such as transitivity that makes the imputation task easier. As the level of missingness is increased, the true positive rate falls more slowly for model 3B with the SLC data in this project than it does for the full model with the Add Health data in C. Wang et al. 2016. As the total percent missingness approachs 35% the median true positive rate of the full Add Health model for schools size 50 to 150 is surpassed by model 3B for the SLC (36.2% versus 36.5%).

3.5 Discussion

For a long time researchers of social networks have been held back by a lack of methods and tools to appropriately handle missing data in their analyses. The most common treatment for missing data was to simply do the analysis with respondents only, discarding the available information about non-respondents including tie nominations of non-respondents by respondents. A growing body of work has begun to provide progress on this front based largely on the development of methods to perform inference for exponential random graphs models with missing data (Handcock and Gile 2010; Robins, Pattison, and Woolcock 2004; Koskinen, Robins, and Pattison 2010; Koskinen et al. 2013). By disseminating these methods and making them available for researchers to use via open source software (Hunter et al. 2008; Caimo and Friel 2014), more and more ERGM analyses are being done with all observations rather than respondents only. Additionally, research from Butts (2003), Peixoto (2018), and others shows that further benefit can be conferred by drawing on information from multiple observations of the dyad. However, despite the methodological developments allowing for the incorporation of multiple noisy measurements into network inference models to increase measurement accuracy, social network data is often not collected with third-party reports.

I contribute to this research by carefully examining how the incorporation of third-party information impacts the performance of model-based imputation. I find that adding information from third-party reports provides a significant boost in imputation accuracy above the state of the art in exponential random graph model-based imputation. This study helps to clarify under what circumstances third-party reports will be most useful. More specifically, I find that the benefit is greater for ties between non-respondents and that the benefit relative to relying on self-reports alone grows with the non-response rate.

Several assumptions and limitations of this study worth considering. First, my results are based on data from a single population. More work is needed to investigate the benefit of third-party information for edge imputation in other contexts. Future evaluations could be done with several already existing cognitive social structure datasets. For example, Siciliano, Yenigun, and Ertan 2012 evaluates an imputation method on five CSS datasets and of these the largest has only 36 actors.

Second, I did not explicitly model informant error in my imputation approach although it likely exists based on the correspondence results of Project 2. This is in large part because models to do so are not yet entirely developed. There is promising work on this front such as Butts 2017 which simultaneously estimates informant accuracy parameters and network structure parameters ($U|man$ in this case). However, much of the previous work has modeled only the overall false positive and false negative error rates (Peixoto 2018; Yenigun, Ertan, and Siciliano 2017; Siciliano, Yenigun, and Ertan 2012) or per informant error rates (Butts 2003). Such models cannot capture structures of respondent bias such as the tendency to overreport own ties (Feld and Carter 2002). There are several works developing Bayesian models of cognitive social structures which attempt to incorporate such hypotheses (Koskinen 2002a, 2002b, 2004; Swartz, Gill, and Muthukumarana 2015; Sosa and Rodriguez 2017; Shao 2018; Sewell 2018). One current complication in the case of complete egocentric network census designs is that third-party reports are dependent on the self-report data and are therefore non-ignorably missing (Butts 2008a). Self-report errors are reflected in the pattern of missingness of third-party reports, and the implications of the missingness structure of third-party reports in an egonet census are unclear. Further research is needed to determine how this might impact measurement error with available imputation methods and whether problems induced by the relationship between self-report error and missingness of third-party report could be ameliorated with new methods that model the dependence.

Third, the SLC self-report survey collected data on three positive relations – hanging out, talking about personal concerns, and talking about school work. I have collapsed this multiplex network data into a single self-report network. While for the purposes of this analysis it was useful to think of the ties between adolescents as representing a single social relation, link prediction research suggests that performance can be improved by simultaneously modeling multiple layers (Tarres-Deulofeu et al. 2019). Indeed, as the field of social network research has benefited from integration of physicists (Freeman 2011; Scott 2011) closer integration with computer science research may yield benefits not only from insights in the subtopics of link prediction and statistical relational learning directly relevant to network reconstruction methods, but also from other areas such as studies of methods for model selection (Valles-Catala et al. 2018) and evaluation. For example, many studies use the area under curve (AUC) of the receiver operating characteristic (ROC) to evaluate predictive performance of missing edge imputation (Clauset, Moore, and Newman 2008; Cranmer and Desmarais 2011; Saul and Filkov 2007), but there is research that shows that precision-recall threshold curves are preferrable to ROC curves due to the class imbalance of edges in sparse network data (Yang, Lichtenwalter, and Chawla 2015). Development of adaptive sampling techniques for social networks might also benefit from research on active learning (Kuwadekar and Neville 2011; Zhuang et al. 2012; Namata et al. 2012; Pfeiffer, Neville, and Bennett 2012; LaRock et al. 2018; Murai Ferreira 2016).

These limitations notwithstanding, one important implication of this project's findings is that harnessing the benefit of third-party reports does not require a complete cognitive social structure design and the associated burden of asking all respondents about the relationship between all dyads in the network. This study demonstrates that there is still a benefit to collecting and incorporating third-party reports even when they are sampled much more sparsely than a complete cognitive social structure, such as with an ego-network census.

More generally, this study suggests researchers conducting social network surveys may wish to consider third-party reports as part of their missing data strategy, especially if non-response is high. Third-party information may not be necessary when response rate and reciprocity are high, but imputation accuracy is reduced as nonresonse goes up. I've shown that third-party reports offer a source of information on edges for which there is no self-report data and help to maintain imputation performance in the presence of significant non-response. Collecting third-party reports could allow researchers to successfully perform analysis with all observations rather than resorting to a respondents only approach in cases where non-response is sufficiently high that model-based imputation performance would be overly degraded when relying on self-report data alone.

NOTES

adams, jimi, and James Moody. 2007. "To tell the truth: Measuring concordance in multiply reported network data." *Social Networks* 29 (1). doi:10.1016/j.socnet.2005.11.009.

Agresti, Alan. 2012. *Categorical Data Analysis.* 3rd ed. John Wiley & Sons.

An, Weihua, and Sam Schramski. 2015. "Analysis of contested reports in exchange networks based on actors' credibility." *Social Networks* 40:25–33.

Argys, Laura M., Daniel I. Rees, and Dominic J. Brewer. 1996. "Detracking America's schools: Equity at zero cost?" *Journal of Policy Analysis and Management* 15 (4): 623–645.

Back, Mitja D., and David A. Kenny. 2010. "The social relations model: How to understand dyadic processes." *Social and Personality Psychology Compass* 4 (10): 855–870.

Bell, David C., Benedetta Belli-McQueen, and Ali Haider. 2007. "Partner naming and forgetting: recall of network members." *Social Networks* 29 (2): 279–299.

Bernard, H. Russell, and Peter D. Killworth. 1977. "Informant accuracy in social network data II." *Human Communication Research* 4 (1): 3–18.

Bernard, H. Russell, Peter D. Killworth, and Lee Sailer. 1979. "Informant accuracy in social network data IV: A comparison of clique-level structure in behavioral and cognitive network data." *Social Networks* 2 (3): 191–218.

———. 1982. "Informant accuracy in social-network data V. An experimental attempt to predict actual communication from recall data." *Social Science Research* 11 (1): 30–66.

Bernard, H. Russell, Peter Killworth, David Kronenfeld, and Lee Sailer. 1984. "The problem of informant accuracy: The validity of retrospective data." *Annual Review of Anthropology* 13 (1): 495–517.

Bill McEvily. 2014. "Do you know My Friend? Attending to the Accuracy of Egocentered Network Data." In *Contemporary Perspectives on Organizational Social Networks,* 40:295–313. Research in the Sociology of Organizations 40. Emerald Group Publishing Limited. Accessed April 20, 2017. doi:10.1108/S0733-558X(2014)0000040015.

Bonacich, Phillip. 1987. "Power and centrality: A family of measures." *American Journal of Sociology* 92 (5): 1170–1182. Accessed April 11, 2017. http://www.journals.uchicago.edu/doi/abs/10.1086/228631.

Bondonio, Daniele. 1998. "Predictors of accuracy in perceiving informal social networks." *Social Networks* 20 (4): 301–330. Accessed April 26, 2017. http://www.sciencedirect.com/science/article/pii/S0378873398000070.

Borgatti, Stephen P., and Martin G. Everett. 2006. "A graph-theoretic perspective on centrality." *Social Networks* 28 (4): 466–484.

Brandes, Ulrik. 2008. "On variants of shortest-path betweenness centrality and their generic computation." *Social Networks* 30, no. 2 (): 136–145. Accessed April 11, 2017. doi:10.1016/j.socnet.2007.11.001.

Brugere, Ivan, Brian Gallagher, and Tanya Y. Berger-Wolf. 2018. "Network structure inference, a survey: Motivations, methods, and applications." *ACM Computing Surveys (CSUR)* 51 (2): 24.

Burt, Ronald S., and William M. Bittner. 1981. "A note on inferences regarding network subgroups." *Social Networks* 3 (1): 71–88.

Butts, Carter T. 2003. "Network inference, error, and informant (in) accuracy: a Bayesian approach." *Social Networks* 25 (2): 103–140.

———. 2008a. "Social network analysis: A methodological introduction." *Asian Journal of Social Psychology* 11 (1): 13–41. Accessed April 12, 2017. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-839X.2007.00241.x/full.

———. 2008b. "Social network analysis with sna." *Journal of Statistical Software* 24 (6): 1–51.

———. 2017. "Baseline Mixture Models for Social Networks." *arXiv e-prints.* https://arxiv.org/abs/1710.02773.

Caimo, Alberto, and Nial Friel. 2014. "Bergm: Bayesian Exponential Random Graphs in R." *Journal of Statistical Software* 61 (2): 1–25.

Cairns, Robert B., Jean-Louis Gariepy, Thomas Kindermann, and M.C. Leung. 1989. "Identifying social clusters in natural settings." *Unpublished manuscript, University of North Carolina at Chapel Hill.*

Cappella, Elise, Jennifer Watling Neal, and Neha Sahu. 2012. “Children’s agreement on classroom social networks: cross-level predictors in urban elementary schools.” *Merrill-Palmer Quarterly (1982-):* 285–313.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2008. *Does your cohort matter? Measuring peer effects in college achievement.* Technical report. National Bureau of Economic Research.

Casciaro, Tiziana. 1998. “Seeing things clearly: Social structure, personality, and accuracy in social network perception.” *Social Networks* 20 (4): 331–351.

Casciaro, Tiziana, Kathleen M. Carley, and David Krackhardt. 1999. “Positive affectivity and accuracy in social network perception.” *Motivation and Emotion* 23 (4): 285–306.

Clauset, Aaron, Cristopher Moore, and Mark E.J. Newman. 2008. “Hierarchical structure and the prediction of missing links in networks.” *Nature* 453 (7191): 98.

Clauset, Aaron, Ellen Tucker, and Matthias Sainz. 2016. *The colorado index of complex networks.* https://icon.colorado.edu/.

Cornell, Dewey G., Peter J. Lovegrove, and Michael W. Baly. 2014. “Invalid survey response patterns among middle school students.” *Psychological Assessment* 26 (1): 277.

Cranmer, Skyler J., and Bruce A. Desmarais. 2011. “Inferential network analysis with exponential random graph models.” *Political Analysis* 19 (1): 66–86.

D’Andrade, Roy G. 1965. “Trait psychology and componential analysis.” *American Anthropologist* 67 (5): 215–228.

———. 1973. “Cultural constructions of reality.” *Cultural illness and health:* 115–27.

———. 1974. “Memory and the assessment of behavior.” *Measurements in the Social Sciences: Theories and Strategies:* 159.

Daniel, João R., Rita R. Silva, and António J. Santos. 2018. “Network Structure Predicts Changes in Perception Accuracy of Social Relationships.” *Frontiers in psychology* 9.

De Nooy, Wouter, Andrej Mrvar, and Vladimir Batagelj. 2018. *Exploratory social network analysis with Pajek: Revised and expanded edition for updated software.* Vol. 46. Cambridge University Press.

Doreian, Patrick, Roman Kapuscinski, David Krackhardt, and Janusz Szczypula. 1996. "A brief history of balance through time." *Journal of Mathematical Sociology* 21 (1-2): 113–131.

Eckles, James E., and Eric G. Stradley. 2012. "A social network analysis of student retention using archival data." *Social Psychology of Education* 15 (2): 165–180.

Faust, Katherine. 2010. "A puzzle concerning triads in social networks: Graph constraints and the triad census." *Social Networks* 32 (3): 221–233.

Feld, Scott L., and William C. Carter. 2002. "Detecting measurement bias in respondent reports of personal networks." *Social Networks* 24 (4): 365–383.

Flashman, Jennifer. 2012. "Academic achievement and its impact on friend dynamics." *Sociology of Education* 85 (1): 61–80.

Foster, Gigi. 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.

Frank, Kenneth A. 1995. "Identifying cohesive subgroups." *Social Networks* 17 (1): 27–56.

———. 1996. "Mapping interactions within and between cohesive subgroups." *Social Networks* 18 (2): 93–119.

Frank, Kenneth A., Chandra Muller, and Anna S. Mueller. 2013. "The Embeddedness of Adolescent Friendship Nominations: The Formation of Social Capital in Emergent Network Structures1." *American Journal of Sociology* 119 (1): 216–253.

Freeman, Linton C. 1978. "Centrality in social networks conceptual clarification." *Social Networks* 1, no. 3 (): 215–239. Accessed April 11, 2017. doi:10.1016/0378-8733(78)90021-7.

———. 1992. "Filling in the blanks: A theory of cognitive categories and the structure of social affiliation." *Social Psychology Quarterly:* 118–127.

———. 2011. "The development of social network analysis–with an emphasis on recent events." Chap. 3 in *The SAGE handbook of social network analysis,* edited by John Scott and Peter J. Carrington, 26–39. SAGE Publications Ltd.

Freeman, Linton C., and A. Kimball Romney. 1987. "Words, deeds and social structure: A preliminary study of the reliability of informants." *Human Organization* 46 (4): 330–334.

Freeman, Linton C., A. Kimball Romney, and Sue C. Freeman. 1987. "Cognitive structure and informant accuracy." *American anthropologist* 89 (2): 310–325.

Friedkin, Noah E. 1983. "Horizons of observability and limits of informal control in organizations." *Social Forces* 62 (1): 54–77.

Friedkin, Noah E. 1991. "Theoretical foundations for centrality measures." *American journal of Sociology* 96 (6): 1478–1504.

Garlaschelli, Diego, and Maria I. Loffredo. 2004. "Patterns of link reciprocity in directed networks." *Physical review letters* 93 (26): 268701.

Gasevic, Dragan, Amal Zouaq, and Robert Janzen. 2013. ""Choose Your Classmates, Your GPA Is at Stake!": The Association of Cross-Class Social Ties and Academic Performance." *American Behavioral Scientist:* 0002764213479362.

Gelman, Andrew. 2008. "Scaling regression inputs by dividing by two standard deviations." *Statistics in medicine* 27 (15): 2865–2873.

Goodreau, Steven M., James A. Kitts, and Martina Morris. 2009. "Birds of a feather, or friend of a friend? using exponential random graph models to investigate adolescent social networks." *Demography* 46 (1): 103–125.

Guimera, Roger, and Marta Sales-Pardo. 2009. "Missing and spurious interactions and the reconstruction of complex networks." *Proceedings of the National Academy of Sciences* 106 (52): 22073–22078.

Handcock, Mark S., and Krista J. Gile. 2010. "Modeling social networks from sampled data." *The Annals of Applied Statistics* 4 (1): 5.

Harris, Kathleen M., Francesca Florey, Joyce Tabor, Peter S. Bearman, Jo Jones, and J. Richard Udry. 2009. *The national longitudinal study of adolescent health: research design.*

Hedeker, Donald, Hakan Demirtas, and Robin J. Mermelstein. 2009. "A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data." *Statistics and its Interface* 2 (4): 391.

Hipp, John R., Cheng Wang, Carter T. Butts, Rupa Jose, and Cynthia M. Lakon. 2015. "Research note: The consequences of different methods for handling missing network data in stochastic actor based models." *Social Networks* 41:56–71.

Hoff, Peter D. 2009. "Multiplicative latent factor models for description and prediction of social networks." *Computational and mathematical organization theory* 15 (4): 261.

Holland, Paul W., and Samuel Leinhardt. 1970. "A Method for Detecting Structure in Sociometric Data." *American Journal of Sociology:* 492–513.

Huisman, Mark. 2009. "Imputation of missing network data: some simple procedures." *Journal of Social Structure* 10 (1): 1–29.

Huisman, Mark, and Robert W. Krause. 2017. "Imputation of missing network data." In *Encyclopedia of Social Network Analysis and Mining,* 2nd ed., edited by Reda Alhajj and Jon Rokne, 1044–1053. Springer.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008. "ergm: A package to fit, simulate and diagnose exponential-family models for networks." *Journal of statistical software* 24 (3): nihpa54860.

Jensen, David, Jennifer Neville, and Brian Gallagher. 2004. "Why collective inference improves relational classification." In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining,* 593–598. ACM.

Johnson, Jeffrey C., and Michael K. Orbach. 2002. "Perceiving the political landscape: ego biases in cognitive political networks." *Social Networks* 24 (3): 291–310.

Kashy, Deborah A., and David A. Kenny. 1990. "Do you know whom you were with a week ago Friday? A re-analysis of the Bernard, Killworth, and Sailer studies." *Social Psychology Quarterly:* 55–61.

Keele, Luke, and David K. Park. 2006. "Difficult choices: an evaluation of heterogeneous choice models." In *2004 Meeting of the American Political Science Association.*

Kilduff, Martin, Craig Crossland, Wenpin Tsai, and David Krackhardt. 2008. "Organizational network perceptions versus reality: A small world after all?" *Organizational Behavior and Human Decision Processes* 107 (1): 15–28.

Killworth, Peter D., and H. Russell Bernard. 1976. "Informant accuracy in social network data." *Human Organization:* 269–286.

———. 1979. "Informant accuracy in social network data III: A comparison of triadic structure in behavioral and cognitive data." *Social Networks* 2 (1): 19–46.

Kim, Myunghwan, and Jure Leskovec. 2011. "The network completion problem: Inferring missing nodes and edges in networks." In *Proceedings of the 2011 SIAM International Conference on Data Mining,* 47–58. SIAM.

Koskinen, Johan H. 2002a. *Bayesian analysis of cognitive social structures with covariates.* Technical report.

Koskinen, Johan H. 2002b. *Bayesian analysis of perceived social networks.* Technical report.

———. 2004. *Model selection for cognitive social structures.* Technical report. Research Report 2004: 3, Department of Statistics, Stockholm University ….

Koskinen, Johan H., Garry L. Robins, and Philippa E. Pattison. 2010. "Analysing exponential random graph (p-star) models with missing data using Bayesian data augmentation." *Statistical Methodology* 7 (3): 366–384.

Koskinen, Johan H., Garry L. Robins, Peng Wang, and Philippa E. Pattison. 2013. "Bayesian analysis for partially observed network data, missing ties, attributes and actors." *Social Networks* 35 (4): 514–527.

Kossinets, Gueorgi. 2006. "Effects of missing data in social networks." *Social Networks* 28 (3): 247–268.

Kossinets, Gueorgi, and Duncan J. Watts. 2006. "Empirical analysis of an evolving social network." *science* 311 (5757): 88–90.

Krackhardt, David. 1987. "Cognitive social structures." *Social Networks* 9 (2): 109–134.

———. 1990. "Assessing the political landscape: Structure, cognition, and power in organizations." *Administrative science quarterly:* 342–369.

———. 1996. "Comment on Burt and Knez's third-party effects on trust." *Rationality and Society* 8 (1): 111–116.

Krackhardt, David, and Martin Kilduff. 1999. "Whether close or far: Perceptions of balance in friendship networks in organizations." *Journal of Personality and Social Psychology* 76 (5): 770–782.

Krause, Robert W., Mark Huisman, Christian Steglich, and Tom A.B. Sniiders. 2018. "Missing network data a comparison of different imputation methods." In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM),* 159–163. IEEE.

Kumbasar, Ece, A. Kimball Rommey, and William H. Batchelder. 1994. "Systematic biases in social perception." *American journal of sociology* 100 (2): 477–505.

Kunegis, J. 2013. "Proceedings of the 22nd International Conference on World Wide Web."

Kuwadekar, Ankit, and Jennifer Neville. 2011. "Relational active learning for joint collective classification models." In *Proceedings of the 28th International Conference on Machine Learning (ICML-11),* 385–392. Citeseer.

LaRock, Timothy, Timothy Sakharov, Sahely Bhadra, and Tina Eliassi-Rad. 2018. "Reducing Network Incompleteness Through Online Learning: A Feasibility Study."

Lee, Francis, and Carter T. Butts. 2018. "Mutual assent or unilateral nomination? A performance comparison of intersection and union rules for integrating self-reports of social relationships." *Social Networks* 55:55–62.

Liben-Nowell, David, and Jon Kleinberg. 2007. "The link-prediction problem for social networks." *Journal of the American society for information science and technology* 58 (7): 1019–1031.

Lomi, Alessandro, Tom A.B. Snijders, Christian E.G. Steglich, and Vanina Jasmine Torló. 2011. "Why are some more peer than others? Evidence from a longitudinal study of social networks and individual academic performance." *Social Science Research* 40 (6): 1506–1520.

Lu, Linyuan, and Tao Zhou. 2011. "Link prediction in complex networks: A survey." *Physica A: statistical mechanics and its applications* 390 (6): 1150–1170.

Malloy, Thomas E., Linda Albright, and Stan Scarpati. 2007. "Awareness of peers' judgments of oneself: Accuracy and process of metaperception." *International Journal of Behavioral Development* 31 (6): 603–610.

Marcum, Christopher Steven, Christine A. Bevc, and Carter T. Butts. 2012. "Mechanisms of control in emergent interorganizational networks." *Policy Studies Journal* 40 (3): 516–546.

Maroulis, Spiro, and Louis M. Gomez. 2008. "Does "connectedness" matter? Evidence from a social network analysis within a small-school reform." *The Teachers College Record* 110 (9): 1901–1929.

Marsden, Peter V. 2005. "Recent developments in network measurement." 00271, *Models and methods in social network analysis* 8:30. Accessed October 27, 2013. http://books.google.com.ezproxy1.lib.asu.edu/books?hl=en&lr=&id=4Ty5xP_KcpAC&oi=fnd&pg=PA8&dq=Recent+developments+in+network+measurement&ots=9LLICdA6H6&sig=GDFVch9MbnR_cOiHpuPAuB1zPPs.

Marsden, Peter V., and Karen E. Campbell. 1984. "Measuring tie strength." *Social forces* 63 (2): 482–501.

Marsden, Peter V., and Karen E. Campbell. 2012. "Reflections on conceptualizing and measuring tie strength." *Social forces* 91 (1): 17–23.

Matzat, Uwe, and Chris Snijders. 2010. "Does the online collection of ego-centered network data reduce data quality? An experimental comparison." *Social Networks* 32 (2): 105–111.

McCarty, Christopher, Peter D. Killworth, and James Rennell. 2007. "Impact of methods for reducing respondent burden on personal network structural measures." *Social Networks* 29 (2): 300–315.

Meijs, Noortje, Antonius H.N. Cillessen, Ron H.J. Scholte, Eliane Segers, and Renske Spijkerman. 2010. "Social intelligence and academic achievement as predictors of adolescent popularity." *Journal of Youth and Adolescence* 39 (1): 62.

Miller, Kurt, Michael I. Jordan, and Thomas L. Griffiths. 2009. "Nonparametric latent feature models for link prediction." In *Advances in neural information processing systems,* 1276–1284.

Moody, James. 2001. "Race, school integration, and friendship segregation in america1." *American Journal of Sociology* 107 (3): 679–716.

Morgan, Stephen L., and Aage B. Sorensen. 1999. "Parental networks, social closure, and mathematics learning: A test of Coleman's social capital explanation of school effects." *American Sociological Review:* 661–681.

Murai Ferreira, Fabricio. 2016. "Applications of Sampling and Estimation on Networks." PhD diss., University of Massachussetts Amherst.

Namata, Galileo, Ben London, Lise Getoor, and Bert Huang. 2012. "Query-driven active surveying for collective classification." In *10th International Workshop on Mining and Learning with Graphs.*

Neal, Jennifer Watling. 2008. """Kracking" the missing data problem: applying Krackhardt's cognitive social structures to school-based social networks." *Sociology of Education* 81 (2): 140–162. Accessed April 26, 2017. http://journals.sagepub.com/doi/abs/10.1177/003804070808100202.

Neal, Jennifer Watling, and Elise Cappella. 2014. "The bright side of positive perceptual bias: Children's estimations of network centrality and aggression." *Aggressive behavior* 40 (2): 140–151.

Neal, Jennifer Watling, Zachary P. Neal, and Elise Cappella. 2016. "Seeing and being seen: predictors of accurate perceptions about classmates' relationships." *Social Networks* 44:1–8.

Newman, Mark E.J. 2018. "Estimating network structure from unreliable measurements." *Physical Review E* 98 (6): 062321.

Oakes, Jeannie, et al. 1990. "Multiplying Inequalities: The Effects of Race, Social Class, and Tracking on Opportunities to Learn Mathematics and Science."

Oguz, Ahmet Tamer. 2014. "Social Network Analysis Using Perceptual Tomography." Master's thesis.

Ouellette, David M. 2008. "Shadows on the Cave Wall: The Cognitive Accuracy of Social Network Perception." PhD diss., Virginia Commonwealth University.

Parker, James D.A., Ronald E. Creque Sr., David L. Barnhart, Jan Irons Harris, Sarah A. Majeski, Laura M. Wood, Barbara J. Bond, and Marjorie J. Hogan. 2004. "Academic achievement in high school: does emotional intelligence matter?" *Personality and individual differences* 37 (7): 1321–1330.

Peixoto, Tiago P. 2014. "The graph-tool python library." *figshare.*

———. 2018. "Reconstructing networks with unknown and heterogeneous errors." *Physical Review X* 8 (4): 041011.

Pfeiffer, Joel J., Jennifer Neville, and Paul Bennett. 2012. "Active sampling of networks." In *Proceedings of the Workshop on Mining and Learning with Graphs (MLG-2012).*

Phillips, Nolan Edward. 2017. "In Case of Emergency, Don't Break Glass: The Emergency Management Organizational Field as a Glassy Regime." PhD diss., UC Irvine.

Pittinsky, Matthew Lyle. 2008. "Smart by (perceived) association: Cognitive social networks and teacher academic judgments." PhD diss., Columbia University.

Ridgers, Nicola D., Laura M. Carter, Gareth Stratton, and Thomas L. McKenzie. 2011. "Examining children's physical activity and play behaviors during school playtime over time." *Health education research* 26 (4): 586–595. Accessed April 25, 2017. http://her.oxfordjournals.org/content/26/4/586.short.

Robins, Garry, Philippa Pattison, and Jodie Woolcock. 2004. "Missing data in networks: exponential random graph (p□) models for networks with non-respondents." *Social Networks* 26 (3): 257–283.

Romney, A. Kimball, and Katherine Faust. 1982. "Predicting the structure of a communications network from recalled data." *Social Networks* 4 (4): 285–304.

Romney, A. Kimball, and Susan C. Weller. 1984. "Predicting informant accuracy from patterns of recall among individuals." *Social Networks* 6 (1): 59–77.

Romney, A. Kimball, Susan C. Weller, and William H. Batchelder. 1986. "Culture as consensus: A theory of culture and informant accuracy." *American anthropologist* 88 (2): 313–338.

Sacerdote, Bruce. 2000. *Peer effects with random assignment: Results for Dartmouth roommates.* Technical report. National bureau of economic research.

Saul, Zachary M., and Vladimir Filkov. 2007. "Exploring biological network structure using exponential random graph models." *Bioinformatics* 23 (19): 2604–2611.

Schaefer, David R., Steven A. Haas, and Nicholas J. Bishop. 2012. "A dynamic model of US adolescents' smoking and friendship networks." *American journal of public health* 102 (6): e12–e18.

Schaefer, David R., Sandra D. Simpkins, Andrea E. Vest, and Chara D. Price. 2011. "The contribution of extracurricular activities to adolescent friendships: New insights through social network analysis." *Developmental psychology* 47 (4): 1141.

Scott, John. 2011. "Social physics and social networks." Chap. 5 in *The SAGE handbook of social network analysis,* edited by John Scott and Peter J Carrington, 55–66. SAGE Publications Ltd.

Sewell, Daniel K. 2018. "Latent space models for network perception data." *Network Science:* 1–20.

Shalizi, Cosma Rohilla, and Andrew C. Thomas. 2011. "Homophily and contagion are generically confounded in observational social network studies." *Sociological methods & research* 40 (2): 211–239.

Shao, Kanghong. 2018. "Statistical Models for Cognitive Social Structures." Master's thesis, UCLA.

Siciliano, Michael D., Deniz Yenigun, and Gunes Ertan. 2012. "Estimating network structure via random sampling: Cognitive social structures and the adaptive threshold method." *Social Networks* 34 (4): 585–600.

Snijders, Tom A.B., Gerhard G. Van de Bunt, and Christian E.G. Steglich. 2010. "Introduction to stochastic actor-based models for network dynamics." *Social Networks* 32 (1): 44–60.

Sosa, Juan, and Abel Rodriguez. 2017. "A Latent Space Model for Cognitive Social Structures Data." *arXiv preprint arXiv:1711.03662*.

Stork, Diana, and William D. Richards. 1992a. "Nonrespondents in communication network studies problems and possibilities." *Group & Organization Management* 17 (2): 193–209.

———. 1992b. "Nonrespondents in communication network studies: Problems and possibilities." *Group & Organization Management* 17 (2): 193–209.

Sudman, Seymour, and Norman M. Bradburn. 1973. "Effects of time and memory factors on response in surveys." *Journal of the American Statistical Association* 68 (344): 805–815.

———. 1974. *Response effects in surveys: A review and synthesis.* 16. Aldine.

———. 1983. "Asking questions: a practical guide to questionnaire design."

Swartz, Tim B., Paramjit S. Gill, and Saman Muthukumarana. 2015. "A Bayesian approach for the analysis of triadic data in cognitive social structures." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 64 (4): 593–610.

Tarres-Deulofeu, Marc, Antonia Godoy-Lorite, Roger Guimera, and Marta Sales-Pardo. 2019. "Tensorial and bipartite block models for link prediction in layered networks and temporal networks." *Physical Review E* 99 (3): 032307.

Taskar, Ben, Pieter Abbeel, Ming-Fai Wong, and Daphne Koller. 2007. "Relational markov networks." *Introduction to statistical relational learning:* 175–200.

Taskar, Ben, Ming-Fai Wong, Pieter Abbeel, and Daphne Koller. 2003. "Link prediction in relational data." In *Advances in neural information processing systems,* 659–666.

Thomas, Scott L. 2000. "Ties that bind: A social network approach to understanding student integration and persistence." *Journal of Higher Education:* 591–615.

Valles-Catala, Toni, Tiago P. Peixoto, Marta Sales-Pardo, and Roger Guimera. 2018. "Consistencies and inconsistencies between model selection and link prediction in networks." *Physical Review E* 97 (6): 062316.

Wang, Cheng, Carter T. Butts, John R. Hipp, Rupa Jose, and Cynthia M. Lakon. 2016. “Multiple imputation for missing edge data: A predictive evaluation method with application to Add Health.” *Social Networks* 45:89–98.

Wang, Peng, BaoWen Xu, YuRong Wu, and XiaoYu Zhou. 2015. “Link prediction in social networks: the state-of-the-art.” *Science China Information Sciences* 58 (1): 1–38.

Wasserman, Stanley, and Katherine Faust. 1994. *Social network analysis: Methods and applications.* Vol. 8. Cambridge university press.

Wentzel, Kathryn R., Carolyn McNamara Barry, and Kathryn A. Caldwell. 2004. “Friendships in Middle School: Influences on Motivation and School Adjustment.” *Journal of educational psychology* 96 (2): 195.

Yang, Yang, Ryan N. Lichtenwalter, and Nitesh V. Chawla. 2015. “Evaluating link prediction methods.” *Knowledge and Information Systems* 45 (3): 751–782.

Yenigun, Deniz, Gunecs Ertan, and Michael Siciliano. 2017. “Omission and commission errors in network cognition and network estimation using ROC curve.” *Social Networks* 50:26–34.

Yonezawa, Susan, Amy Stuart Wells, and Irene Serna. 2002. “Choosing tracks: “Freedom of choice” in detracking schools.” *American Educational Research Journal* 39 (1): 37–67.

Zhang, Xuhong, and Carter T. Butts. 2017. “Activity correlation spectroscopy: a novel method for inferring social relationships from activity data.” *Social Network Analysis and Mining* 7 (1): 1.

Zhuang, Honglei, Jie Tang, Wenbin Tang, Tiancheng Lou, Alvin Chin, and Xia Wang. 2012. “Actively learning to infer social ties.” *Data Mining and Knowledge Discovery* 25 (2): 270–297.

Zimmerman, David J. 2003. “Peer effects in academic outcomes: Evidence from a natural experiment.” *Review of Economics and Statistics* 85 (1): 9–23.

Znidarsic, Anja, Anuska Ferligoj, and Patrick Doreian. 2012. “Non-response in social networks: The impact of different non-response treatments on the stability of blockmodels.” *Social Networks* 34 (4): 438–450.

APPENDIX A

SMALL LEARNING COMMUNITY DATA

A.1 Multiple Choice

The multiple choice questions were

1. How often did your teachers talk to your parents or another adult living with you this year?
   a) Never
   b) 1 or 2 times
   c) 3 - 5 times
   d) More than 5 times
   e) I don't know
2. On a typical day this past year, how much time did you spend outside of class studying or doing homework?
   a) None
   b) Less than 30 minutes
   c) 30 - 60 minutes
   d) 1 - 2 hours
   e) More than 2 hours
3. This past school year, how often have you had a conversation with another student about going to college?
   a) Never
   b) Once or twice all year
   c) About once or twice a month
   d) About once or twice a week
   e) Almost every day
4. This past school year, how often have you helped another student with their homework, or something you were doing in class that the other student did not understand?
   a) Never
   b) Once or twice all year
   c) About once or twice a month
   d) About once or twice a week
   e) Almost every day
5. This past school year, how often have you encouraged another student to work hard at school?
   a) Never
   b) Once or twice all year
   c) About once or twice a month

d) About once or twice a week
e) Almost every day

6. What do you plan on doing after graduation?
    a) Getting a job
    b) Going to college
    c) Joining the military
    d) I'm not sure yet

A.2 Social Network Survey Questions

A.2.1 Self-report Survey Questions

There were minor deviations in the wording of the social network survey questions. The exact text of each name generator question used is shown in Table 8. The exact text of both sub-questions about each nomination is shown in Table 9.

A.2.2 Third Party Links Survey Question

The interface for the third party links survey is shown in Figure 2. The text of the third party links survey question is shown in Table 10.

1A	Please type in the first and last names of the 7 people you discuss *school work* with the most. The people you list can be students or adults. It is fine to enter less than 7 names if you like.
2A	Please type in the first and last names of the people you discuss *personal and private concerns or worries* with the most. The people you list can be students or adults. It is fine to enter less than 7 names if you like.
3A	Please type in the first and last names of the people you *hang-out* with the most. The people you list can be students or adults. It is fine to enter less than 7 names if you like.
4A	Please type in the first and last names of the people you *avoid*, or would rather not spend time with. The people you list can be students or adults. It is fine to enter less than 7 names if you like.
1B	Please type in the first and last names of the people you discuss *school work* with the most. It is fine to enter fewer than 7 names if you like.
2B	Please type in the first and last names of the people you discuss *personal and private concerns or worries* with the most. It is fine to enter less than 7 names if you like.
3B	Please type in the first and last names of the people you *hang-out* with the most. It is fine to enter fewer than 7 names if you like.
4B	Please type in the first and last names of the people who make it most difficult for you to do your school work. It is fine to enter fewer than 7 names if you like.
1C	Please type in the first and last names of the people you discussed *school work* with the most this year. It is fine to enter fewer than 7 names if you like.
2C	Please type in the first and last names of the people you discussed *personal and private concerns or worries* with the most this year. It is fine to enter less than 7 names if you like.
3C	Please type in the first and last names of the people you *hung-out* with the most this year. It is fine to enter fewer than 7 names if you like.
4C	Please type in the first and last names of the people who made it most difficult for you to do your school work this year. It is fine to enter fewer than 7 names if you like.

Table 8. Exact questions used in surveys

Note: Questions 1A to 4A were used in wave 1. Questions 1B to 4B were used in wave 2 and for SLC-A Cohort 1 in wave 3. Questions 1C to 4C were used in wave 3 for Cohort 3.

For each person, please tell us whether that person is:

- A student in this school
- An adult in this school
- A student outside of this school
- An adult outside of this school

For each person, please also tell us if the person is a relative of yours. For example, if the person is your cousin you would choose the option "Relative - other".

- Not a relative
- Relative - parent
- Relative - sibling
- Relative - other

Table 9. Person type and relation type questions

Note: These questions were asked for each nomination.

This is the last question. In this question, we want to know what you know about the relationship between the people you listed on the previous page. At the bottom of this page is a table with the names of all the names you entered. Please read the directions below, then fill out your table accordingly.
We want you to select "Often" when you think that two people speak often.
We want you to select "Some" when you think that two people speak sometimes.
We want you to select "Rare" when you think that two people speak rarely, or not at all.
We want you to select "Don't know" when you do not know whether or not the two people speak to each other.
[...]
Here is your table with the names of the people you listed on the previous page on the diagonal. Please use the pull-down menus to indicate your view of their relationship. If you want to change or add names, hit your browser's "BACK" button, make the changes, and then return here to complete this last question.

Table 10. Question text for third party links survey

APPENDIX B

ADDITIONAL SELECTION ANALYSES

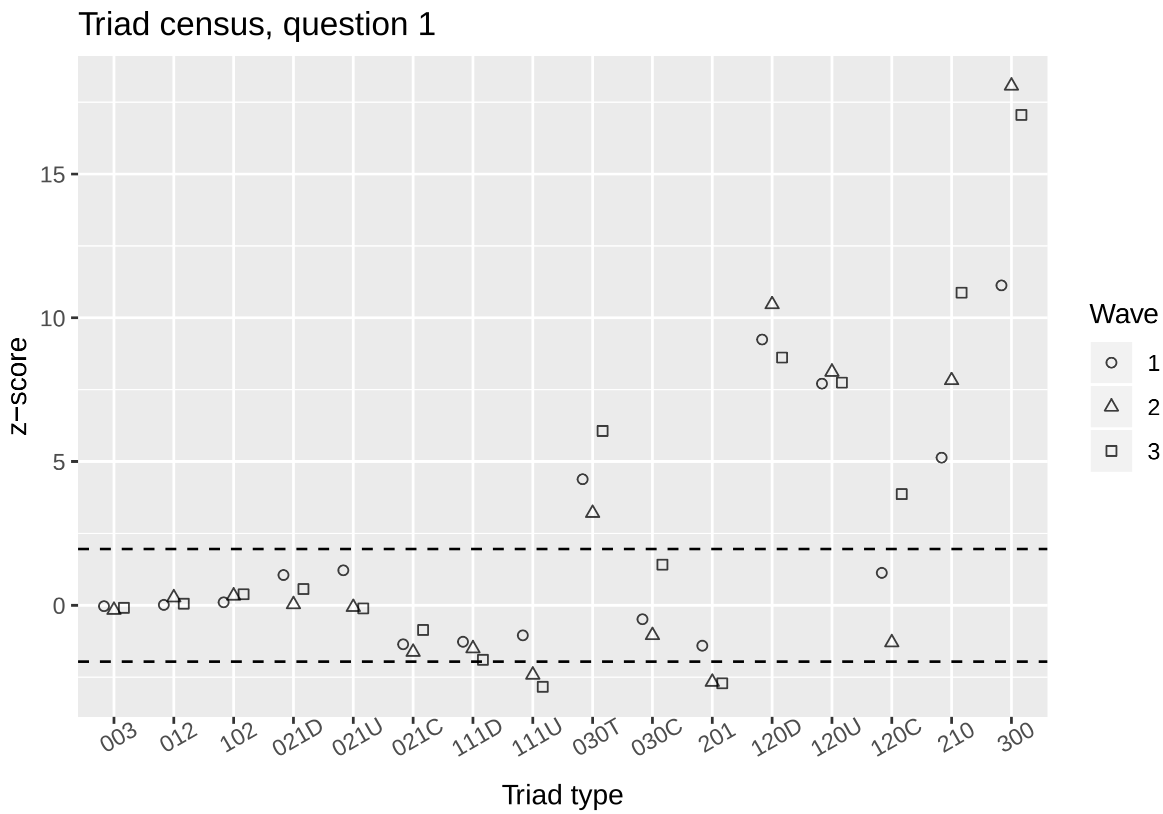


Figure 13. Triad census of the SLC-1 question 1 networks

B.1 Triad Censuses for Each Name Generator

Figures 13, 14, and 15 present the z-scores of the triad censuses for networks corresponding to each name generator question. Introduced in Section 1.3.6.

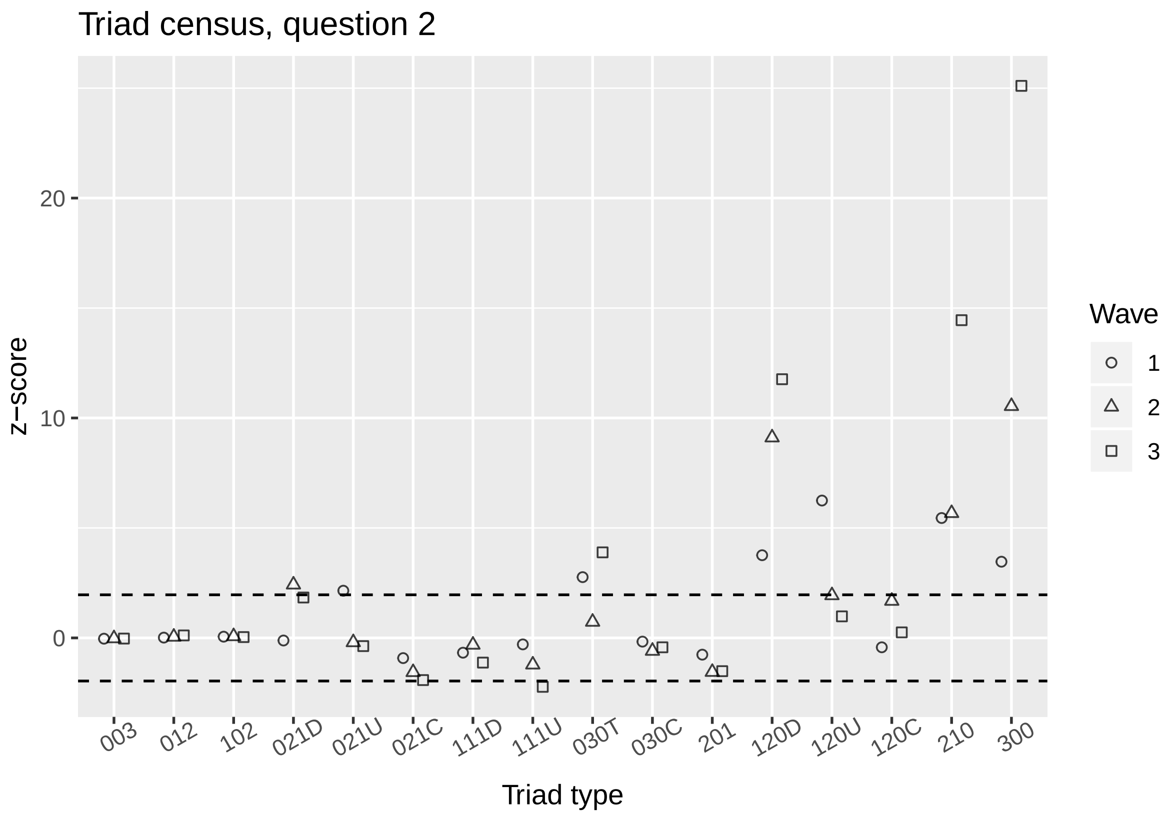


Figure 14. Triad census of the SLC-1 question 2 networks

B.2 Hierarchy in the Small Learning Community

B.2.1 Background

Frank, Muller, & Mueller (2013) find evidence that status hierarchies are a factor in student friendship formation and that the tendency towards hierarchy is reduced within local positions.

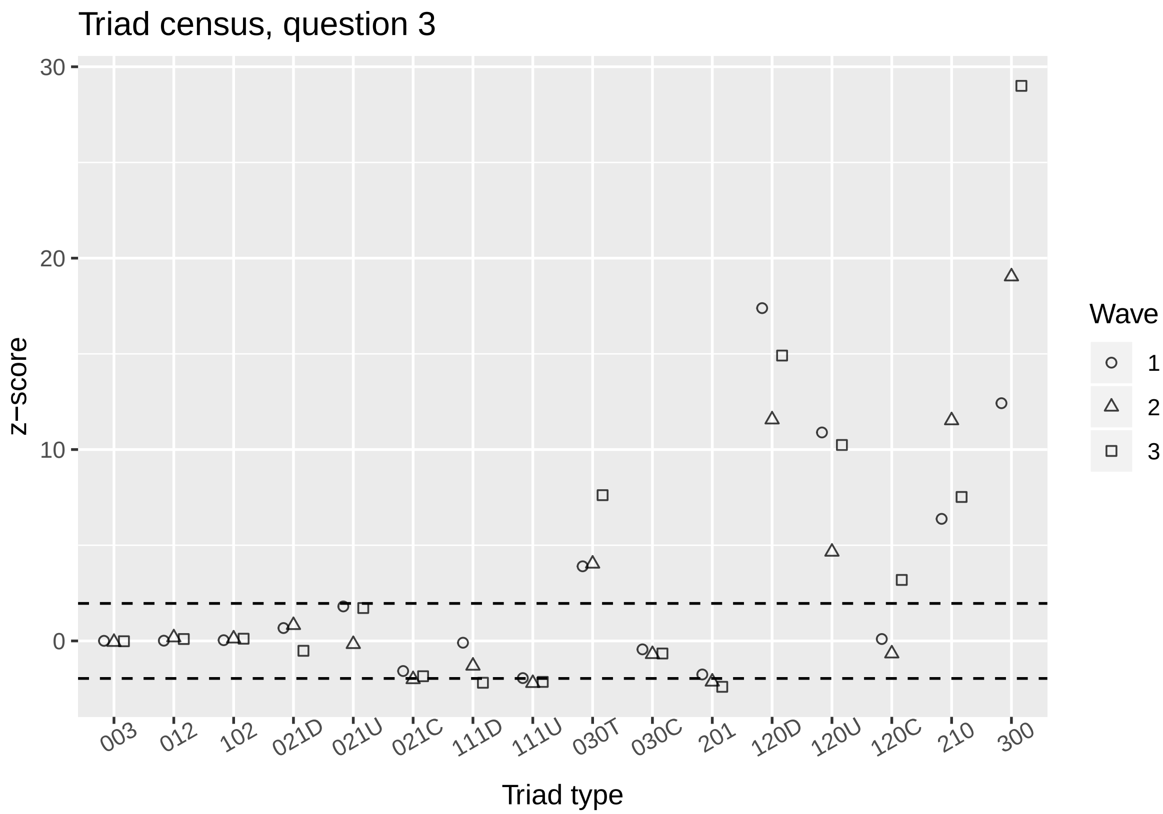


Figure 15. Triad census of the SLC-1 question 3 networks

B.2.2 Hypothesis

Proponents of small schools reform claim that SLCs promote a more cohesive social environment and that this is beneficial for norm enforcement and keeping students engaged. This might imply that the social structure will be non-hierarchical.

Hypothesis 4 *The social structure in the SLC will be non-hierarchical.*

B.3 Tendency to Complete Transitivity Versus Cycles

B.3.1 Method

One way to measure the level of hierarchy is via the tendency to complete transitivity versus cycles Frank, Muller, and Mueller 2013. We first take all new, unreciprocated ties in waves 2 and 3. We stack these and perform a network logistic regression on them including a term for the potential to complete transitivity versus cycles in the previous wave. This is equal to the number of 2-paths from ego to alter minus the number of 2-paths from alter to ego.

B.3.2 Results

Results are presented in Table 11. Although strong social cohesion might imply little hierarchy (Hypothesis 4), we do find a weakly significant coefficient on potential to complete transitivity versus cycles. If we interact the potential to complete transitivity versus cycles with GPA similarity, the effect becomes more significant.

B.4 Mutual Ties

This section replicates analyses from Project 1 with a different network definition where only mutual ties are considered.

	Hierarchy 1	Hierarchy 2
Intercept	0.38	0.42
	(0.78)	(0.80)
Same ZIP	-0.09	-0.13
	(0.42)	(0.43)
Same gender	0.53	$0.74^{\cdot}$
	(0.39)	(0.41)
GPA similarity	-0.99	-1.20
	(1.02)	(1.08)
PCTVC	$0.53^{\cdot}$	-2.70^{*}
	(0.28)	(1.34)
PCTVC * (GPA similarity)		4.73^{*}
		(1.86)
AIC	179.77	170.49
BIC	193.99	187.56
Log Likelihood	-84.89	-79.25
Deviance	169.77	158.49
Num. obs.	127	127

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 11. Potential to complete transitivity versus cycles

Note: Logistic regression on new, unreciprocated ties with term for potential to complete transitivity versus cycles (PCTVC)

	Model 1	Model 2	Model 3
edges	-5.68^{***}	-6.22^{***}	-6.63^{***}
	(0.77)	(0.62)	(0.69)
gwdegree	1.56^{**}	1.27^{**}	1.09^{*}
	(0.51)	(0.42)	(0.45)
nodefactor.Female.TRUE	$-1.15^{\cdot}$	-0.74	0.33
	(0.62)	(0.55)	(0.60)
nodematch.Female	2.24^{**}	1.56^{**}	0.85
	(0.68)	(0.55)	(0.62)
nodematch.ZIP	0.30	0.27	0.11
	(0.32)	(0.22)	(0.23)
edgecov.GPASIM	-1.49	-0.22	1.87
	(1.44)	(1.18)	(1.25)
edgecov.FEMALEEITHER * GPASIM	3.35^{*}	2.16	-0.54
	(1.66)	(1.38)	(1.42)
gwesp.fixed.0.1	1.17^{***}	1.65^{***}	1.40^{***}
	(0.23)	(0.21)	(0.20)
AIC	481.50	671.74	677.54
BIC	526.12	718.95	723.85
Log Likelihood	-232.75	-327.87	-330.77

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 12. Cross-sectional exponential random graph models of mutual ties

Note: Standard errors in parentheses.

	Model 1	Model 2
edges	-10.83^{***}	-8.89^{***}
	(1.28)	(0.84)
gwdegree	1.19^{***}	1.19^{***}
	(0.34)	(0.35)
nodefactor.MSTA.TRUE	0.14	0.36^{*}
	(0.13)	(0.18)
nodefactor.Female.TRUE	$-1.20^{\cdot}$	$-0.82^{\cdot}$
	(0.65)	(0.49)
nodecov.GPA	0.06	-0.02
	(0.09)	(0.08)
nodematch.MSTA	4.13^{***}	2.33^{***}
	(0.96)	(0.45)
nodematch.ZIP	0.58^{**}	0.32
	(0.23)	(0.20)
nodematch.Female	2.46^{***}	1.75^{***}
	(0.67)	(0.51)
edgecov.GPASIM	-1.19	-0.66
	(1.37)	(1.04)
edgecov.FEMALEEITHER * GPASIM	3.35^{*}	$2.37^{\cdot}$
	(1.60)	(1.25)
gwesp.fixed.0.1	1.24^{***}	1.60^{***}
	(0.19)	(0.18)
AIC	789.48	926.16
BIC	864.53	1001.77
Log Likelihood	-383.74	-452.08

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 13. Cross-sectional exponential random graph models of mutual ties

Note: Control group included. Standard errors in parentheses.

	Model 1	Model 2
edges	$-8.48\ (1.24)^{***}$	$-5.08\ (0.81)^{***}$
gwdegree	$0.13\ (0.49)$	$0.16\ (0.47)$
edgecov.mat1	$4.39\ (2.12)^{*}$	$2.59\ (1.72)$
edgecov.ntrans1	$2.30\ (0.45)^{***}$	$1.14\ (0.46)^{*}$
edgecov.mat1 * ntrans1	$-1.92\ (0.69)^{**}$	$-0.54\ (0.52)$
nodematch.Female	$1.05\ (0.43)^{*}$	$0.81\ (0.32)^{*}$
nodematch.ZIP	$0.46\ (0.38)$	$-0.46\ (0.37)$
edgecov.gpa_sim	$4.19\ (1.44)^{**}$	$0.81\ (0.97)$
edgecov.femaleEither * gpa_sim	$0.27\ (0.50)$	
edgecov.mat1 * gpa_sim	$0.04\ (2.57)$	$1.29\ (2.12)$
AIC	304.73	428.11
BIC	357.76	478.01
Log Likelihood	-142.36	-205.05

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, $^{\cdot}p < 0.1$

Table 14. Longitudinal exponential random graph models of mutual ties

Note: Standard errors in parentheses.

APPENDIX C

ADDITIONAL NETWORK MEASUREMENT ANALYSES

C.1 Correspondence Analysis

The self-report networks in our data were summarized by combining information from the same respondent across three relational questions. To summarize third-party reports, information is drawn from multiple students responding to one relational question. As the data was collected with a complete ego network design, most third-party evaluations will be missing. For each dyad e_{ij} third-party reports are summarized with a consensus model which is a function of third-party v_k evaluations of the relation between v_i and v_j for all k.

Such a function might be one when the proportion of the third-party evaluations equal or above an individual threshold θ_I is above the proportion threshold θ_P. This is referred to as the threshold consensus model. Neal (2008) defines a majority rule with $\theta_P = 0.5$, an average rule with θ_P equal to the average proportion of tie nominations, and a binomial rule with θ_P set by the critical value for the binomial distribution with $\alpha = 0.05$. Unless otherwise specified, $\theta_I = \texttt{often}$ and $\theta_P = \frac{2}{3}$.

A direct comparison of networks constructed from the self-report data and third-party reports is made by looking at the presence and absence of ties. Contingency tables and graph correlation are used to compare these measures.

One simple way to summaries a comparison of two categorical measures is with a contingency table. For different network constructions under the same network boundary definition we can compare how they categorize each dyad. For non-directed, unweighted networks there is a simple 2×2 cross-tabulation of the distribution of tie presence and absence in the two networks being compared. To understand the prevalence of type I and type II errors when comparing the third-party report consensus network to the self-report network, the contingency table is reported in Table 15.

	Self-report Relationships		
Ego Net Reported Relationships	Tie Present	Ties Absent	
Ties Present	246	365	611
Ties Absent	280	5998	6278
	526	6363	

Table 15. Contingency table

The correlation of tie presence between the two networks can also be measured. The product-moment correlation between graphs G and H is defined as

$$cor(G, H) = \frac{cov(G, H)}{\sqrt{cov(G, G)cov(H, H)}}$$

where the graph covariance is defined as

$$cov(G, H) = \frac{1}{\binom{|V|}{2}} \sum_{\{i,j\}} (A_{ij}^G - \mu_G)(A_{ij}^H - \mu_H)$$

and A^G is the adjacency matrix of G. Classical null hypothesis testing of the correlation coefficient is not recommended because the dyads are not independent. An alternative method is the Quadratic Assignment Procedure (QAP) which is a permutation test using a Monte Carlo simulation where the rows and columns are randomly reassigned and the empirical distribution of the graph-level statistic is observed. The correlation between self-report and third-party reports is explored using a QAP test of graph correlation. A QAP test of the product-moment correlation between the self-report network and the third-party consensus network was run. In the 10,000 replications there was no case where the correlation with the permuted network was higher than the test value, which was 0.38.

C.1.1 Correlation of Node-level Indices

Usually the interest in collecting social network data is not specifically in the network itself, but in a measure or model derived from the network. Therefore, to understand the implications of network measure insofar as it relates to the constructs of interest we can look at how robust those constructs are to different methods of measurement. For this reason, in addition to direct comparison between third-party and self-report measurements I also compare the correlation of node-level network statistics between networks constructed. Four node-level network statistics that are often used as network regression covariates are compared: betweenness centrality, degree centrality, eigenvector centrality, and ego density. These correlations and bootstrapped confidence intervals are reported in Table 16.

The position of an actor within a social network and how position relates to social processes that play out on the network is of theoretical importance. Node-level indices summarize information about a node's position in the network. Social theories may posit a relation between node-level indices and other variables of interest. For example, it is hypothesized that if a person's connections are well connected to one another, that is if there is network closure, then the person will experience stronger norm enforcement. When comparing network measurements we may wish to compare how well such node-level indices are preserved.

Degree and betweenness centrality were defined in the previous section. A node has high eigenvector centrality if it is connected to other highly central nodes. Eigenvector centrality is calculated by solving the eigenvector equation

$$A^G x = \lambda x$$

Network measure	τ (95% CI)
Betweenness centrality	0.32 (0.17, 0.45)
Degree centrality	0.35 (0.18, 0.50)
Eigenvector centrality	0.41 (0.30, 0.52)
Ego density	0.12 (-0.04, 0.29)

Table 16. Network measure correlations

Correlations are Kendall rank correlation coefficients with bootstrapped 95% CIs.

for the largest eigenvalue λ and corresponding eigenvector x (Bonacich 1987). This implies that all entries in the eigenvector x are nonnegative. The normalized value $\frac{x_i}{\sum_{j=1}^{N} x_j}$ is the centrality score for node i .

Egocentric network density is a measure of network closure. It is calculated by measuring the density of the ego-network. The actual number of ties between an actor's alters is divided by the number of possible ties between them.

C.1.2 Network Logistic Regression

Beyond the direct correlation of the derived networks we can use a network logistic regression to gain a more detailed understanding of the underlying relationship between the network measurements. This can be expressed as

$$\log \frac{p(A^G_{ij} = 1)}{1 - p(A^G_{ij} = 1)} = \beta_0 + \beta_1 x_1(ij) + ... + \beta_m x_m(ij)$$

Where A^G is the adjacency matrix of the dependent network and where $x_k(ij)$ are any dyadic measure. For example, we may look at the how the self-report network is predicted by the different types of responses in the third-party reports. Again there is

Coefficient	Estimate	Exp(b)	Pr(<=b)	Pr(>=b)	Pr(>=|b|)
(intercept)	-3.33504	0.03561	0.116	0.884	0.116
OFTEN	0.75408	2.12565	0.883	0.117	0.247
SOME	-0.70668	0.49327	0.142	0.858	0.293
RARE	-1.57788	0.20641	0.015	0.985	0.043
DONTKNOW	-0.51255	0.59896	0.238	0.762	0.456
OBS > 0	1.53717	4.65140	0.992	0.008	0.022
OBS	0.29398	1.34176	0.766	0.234	0.481

Table 17. Network logistic regression

the problem of network autocorrelation, so Dekker Semi-Partialling MRQAP is used for hypothesis testing.

To further compare the networks, a network logistic regression was used. The dependent variable is tie presence in the self-report network. Given the complete ego-network design, whether there was a third-party evaluator corresponds to whether there was a student that nominated both members of the dyad as alters in the self-report. Together with the clustering and triadic closure found in the network, this means that third-party reports are non-ignorably missing. Therefore the number (OBS) and presence (OBS > 0) of third-party evaluators are included as independent variables. I also include as independent variables the number of evaluators that rated the dyad speaking frequency in each category: OFTEN, SOME, RARE, and DONTKNOW. Results are reported in Table 17. Having at least one third-party evaluator and the number of evaluators that rated the speaking frequency as rare had the largest effect sizes and smallest p-values. The fit indicates that having at least one third-party evaluator captures much of the information about the possible presence of a tie due to the strength of the network clustering.

Table 18. Cumulative link models of third party evaluations with alternate centrality measures

	Model 2.3		Model 2.4	
	Estimate	p-value	Estimate	p-value
Scale Submodel				
Mean KF proximity to alters	-0.287	0.010	-0.320	0.001
Indegree	-0.089	0.385		
Outdegree			-0.179	0.118
Male	0.175	0.088	0.232	0.037
Test scores	-0.259	0.008	-0.307	0.002
Location Submodel				
Alters nominations sum	0.279	0.000	0.288	0.000
Alters KF proximity	0.563	0.000	0.579	0.000
Mean KF proximity to alters	0.135	0.332	0.097	0.483
Percent outside noms	-0.499	0.026	-0.492	0.028

C.2 Ordinal Regression with Dispersion Submodel

C.2.1 Measures of Centrality

Table 18 contains the results for Models 2.3 and 2.4 that respectively have outdegree and indegree centrality in the scale submodel.

C.2.2 Measures of Social Proximity

Table 19 contains the results for Models 2.1.2 and 2.2.2 that replace the KliqueFinder proximity measure with a proximity in a latent social space derived by applying nonmetric multidimensional scaling to the geodesic distance. Table 20 contains the results for Models 2.1.3 and 2.2.3 that replace the KliqueFinder proximity between

Table 19. Cumulative link models of third party evaluations with MDS proximity

	Model 2.1.2		Model 2.2.2	
	Estimate	p-value	Estimate	p-value
Scale Submodel				
Mean MDS proximity to alters	-0.336	0.000	-0.292	0.003
Betweenness	-0.205	0.029		
LAS degree			-0.260	0.003
Male	0.211	0.045	0.191	0.063
Test scores	-0.420	0.000	-0.389	0.000
Location Submodel				
Alters nominations sum	0.332	0.000	0.312	0.000
Alters MDS proximity	0.315	0.012	0.311	0.012
Mean MDS proximity to alters	0.262	0.071	0.249	0.077
Percent outside noms	-0.360	0.108	-0.341	0.123

alters with a binary indicator for whether the alters are in the same KliqueFinder subgroup and replace the average KliqueFinder proximity between ego and alters with a binary indicator for whether the ego is in the same KliqueFinder subgroup with one or both of the alters.

C.3 Don't Know Analysis

Social proximity is negatively correlated with "don't know" responses in third-party evaluations. A mixed effects logistic regression of "don't know" response with a random intercept on the evaluator and fixed effects on the social proximity of the alters and average social proximity from the evaluator to alters finds significant negative coefficients on both proximity effects ($p < 0.05$). Boys are more likely than girls to say don't know. High degree actors are less likely to say don't know. Students with high test scores are more likely to say don't know. Does this contradict hypothesis that high

Table 20. Cumulative link models of third party evaluations with same subgroup proximity

	Model 2.1.3		Model 2.2.3	
	Estimate	p-value	Estimate	p-value
Scale Submodel				
Either alter in ego subgroup	-0.270	0.010	-0.232	0.029
Betweenness	-0.218	0.014		
LAS degree			-0.277	0.002
Male	0.141	0.186	0.143	0.173
Test scores	-0.231	0.020	-0.219	0.027
Location Submodel				
Alters nominations sum	0.302	0.000	0.281	0.000
Alters in same subgroup	0.559	0.000	0.543	0.000
Either alter in ego subgroup	0.041	0.695	0.060	0.559
Percent outside noms	-0.385	0.062	-0.372	0.068

achieving students have more knowledge of peer relationships? Alternatively, does it indicate that they are responding more carefully or is it a Dunning Kruger effect?