# Social Networks:
# Collection, Uncertainty, and Applications to College Student Alcohol Use

by

Alison R. Walsh

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Epidemiological Science)
in the University of Michigan
2019

Doctoral Committee:

Associate Professor Marisa Eisenberg, Chair
Professor Allison Aiello, University of North Carolina
Professor Daniel Eisenberg
Associate Professor Rafael Meza

Alison R. Walsh

walshar@umich.edu

ORCID iD: 0000-0002-6123-5730

# Dedication

This dissertation is dedicated to my sister, Caitlin Walsh, who was the real scholar in the family.

# Acknowledgements

First and foremost, I would like to thank Dr. Marisa Eisenberg for her incredible academic and professional mentorship and guidance, as well as her friendship, during my time in the doctoral program. I would also like to thank Dr. Allison Aiello for being such a dedicated de facto co-mentor and invaluable resource, both at UM and from afar. Marisa and Allison's willingness to let me explore an ever-changing, wide-range of research topics, and their confidence in my work and ability has been an important part of my development as an independent researcher. I am especially thankful for their support and patience during and after my leave of absence for cancer treatment.

I am also incredibly grateful to my other committee members, Drs. Rafael Meza and Daniel Eisenberg, for their thoughtful feedback and questions during my dissertation process. I must also acknowledge the academic and administrative support from the Epi department, including that from other students, faculty, and staff.

I am forever indebted to my family and friends for their support during these last 6 years. Most importantly, I am thankful to my parents for their unwavering support and encouragement. I am also grateful for my extended family's support, especially Dr. Patricia Coogan, without whom I never would have considered pursuing epidemiology. I also must thank all of my friends for being inspiring and remarkable human beings, particularly Lindsay Ciechanski, Josh Dorner, Dr. Velma Lopez, Dr. Sonia Hegde, Scott Hamm, and Dr. Joe Burnett.

Finally, this dissertation was financially supported by the University of Michigan (School of Public Health, Department of Epidemiology, and Rackham Graduate School), the Centers for Disease Control and Prevention (under grant U01CK000185), and the National Institute of General Medical Sciences (under MIDAS grant U01GM110712). I would also like to thank the

University of Michigan School of Public Health, Rackham Graduate School, and the Center for the Education of Women for their generous financial support during my leave of absence.

# Preface

Chapter 2 is currently in the process of being edited for resubmission with the following author list: Alison R. Walsh, Marisa C. Eisenberg*, Allison E. Aiello* (*Co-senior authors)

A version of Table 3.5 was previously published in: Aiello A, Simanek A, Eisenberg M, et al. Design and methods of a social network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial My Citations. *Epidemics*. 2016;15:38–55.

# Table of Contents

# List of Tables

# List of Figures

# Abstract

Behavior, attitudes, and infection can transmit across networks of contacts via social mixing, making network analysis methods a key tool in social and infectious disease epidemiology. Through analysis of the simultaneous processes that influence and shape individuals and networks, we can better understand how to collect social network data, incorporate human behavior and its collective idiosyncrasies into models and statistics as uncertainty, and thus improve the veracity of our conclusions. Using data from a longitudinal social network study of undergraduate students, this dissertation aims to: 1) examine how social structures and contact patterns shape alcohol consumption and use in undergraduate students; 2) evaluate the strengths and limitations of different methods of measuring social contact networks; and 3) develop methods to quantify network uncertainty and hypothesis testing for trait assortativity.

First, we applied social network analysis methods to two undergraduate student social networks, investigating network correlates of alcohol consumption, identifying numerous, consistent associations between alcohol use and social position in this population. Specifically, network position, alcohol exposures, and relationship strength were associated with individual alcohol use, suggesting complex relationships between drinking and network topology, as well as proximity to alcohol use. Overall, this chapter adds to the body of evidence of significant relationships between network structure, social position, and alcohol consumption.

Next, we systematically compared two social network measurement methods with varying levels of granularity in order understand the unique utility of self-report vs. sensor contact data, as well as trends in data quality and quantity over time. Networks were compared across and within each measurement method, using overall network structure, dyad, and node characteristics. We found few network similarities between measurement methods, suggesting that neither empirical network measurement method are complete representations of the underlying "true" social

network. These analyses highlight the impact that network measurement can have on empirical network findings and suggest that researchers should carefully consider which collection method, or combination of methods, could provide them with the highest quality data needed to answer their research questions.

Finally, we outlined and defined multiple assortativity sensitivity analyses, uncertainty quantification approaches, and null model-hypothesis testing procedures and applied these methods to a measured social network of undergraduate students. These investigations showed that uncertainty and biases of attribute assortativity may be predictable, given a defined amount and type of data error.  Generally, results of these analyses show the potential impacts that data quality, measurement error, and the measured network can have on observed assortativity. We suggest that it be standard practice to conduct and present assortativity sensitivity analyses, and to hypothesize possible confounding or bias related to network data quality and completeness.

In toto, this dissertation describes and extensively explores social networks of undergraduate students. We investigated relationships between a risky health behavior of public health importance and network features, as well as how network analysis results using observed networks are reliant on the network measurement method and the types and amounts of data uncertainty and error present. These projects have generated new results and insights into alcohol use and social networks in a college setting, compared empirical social network observations between a traditional and novel instrument, and developed a suite of analytical social network tools. Importantly, the novel methods we defined and implemented in this dissertation provide a framework with which to evaluate network uncertainty, robustness, and hypotheses.

# Chapter 1
# Introduction

## 1.1 Social Network Analysis

As the world becomes more interconnected and less compartmentalized, there are growing dependencies between political, social, and ecological factors, creating a complex environment that significantly and materially impacts individual behavior and health. Network science (the study of how connections and feedbacks between objects or individuals shape actions and vice versa) has been used across disciplines but is particularly important in public health. Methods for analyzing social, or contact, networks are increasingly important tools in social and infectious disease epidemiology (1). Behavior, attitudes, and infection can transmit across networks of contacts via social mixing, making network analysis methods a key tool in social and infectious disease epidemiology. Through analysis of the simultaneous processes that influence and shape individuals and networks, we can better understand how and why behaviors are adopted by individuals and populations. Similarly, elucidation of social networks can clarify infection transmission patterns and inform mathematical models and disease spread predictions. Collectively, social network analysis in public health aims to describe the processes of selection and influence (i.e., transmission). Network influence is the systematic spread of information, ideas, and/or pathogens across the network topology. Research under this umbrella identifies how network structure and properties may facilitate or impede different types of transmission across the network. Network selection is the reverse process, where individual identities and traits are associated with network connectivity and structure. In all likelihood, these are dynamic, bidirectional processes (2–9).

## 1.2 Applications of Social Network Analysis in Public Health

Social network analysis also provides a framework in which to investigate interdependent and circular mechanisms that drive both communicable and non-communicable disease patterns within and across communities. Network analysis is unique in its ability to integrate traditionally defined determinants of health with social, behavioral, and structural factors in order to holistically model health and wellbeing. Granovetter, a sociologist, was one of the first researchers to posit how structures and patterns in realistic social network models, which include both strong and weak connections, could impact human behavior (10), and in the subsequent years, network methods have increasingly been adopted by the fields of infectious and social epidemiology to investigate network associations with a multitude of health behaviors (and their outcomes).

A challenge in infectious disease transmission research is identifying and quantifying how pathogen traits (e.g., mode of transmission, virulence) interact with behavioral, social, and environmental factors that facilitate or restrict transmission between individuals and across populations (11). Most infectious disease modeling and transmission studies rely on homogeneous mixing assumptions (and those that do incorporate heterogeneous contact patterns typically do so with theoretical networks) (12), but it is unclear how closely these resemble real-world contact patterns and thus their results may not accurately describe real-world transmission. Given the importance of predictive transmission modeling for infectious disease prevention and epidemic preparedness, developing more realistic models and parameterizations are crucial areas of public health research (13). Infectious disease dynamics are dependent on a number of factors, including a population's social network structure (1). Indeed, the feedback loops between disease dynamics, individual illness, and behaviors (e.g., isolation, intervention uptake) are likely key components of population disease dynamics and should not be overlooked (14). Social mixing, in particular, can modify and/or mediate the impact of individual factors on population dynamics, particularly for close-contact transmissions (e.g., droplet, direct transmission, and even some forms of indirect transmission) (13). Research has continued to show the relevance of contact patterns and network structure in transmission dynamics and epidemic growth. For instance, the shape of a network's degree distribution can influence an outbreak's basic reproduction number ($R_0$); fat-tailed distributions, wherein a large number of people have a high degree, has been shown to result in a higher than expected $R_0$ (15–17). Clustering and sub-groups in networks, such as communities, small-world degree distributions, node groupings, and modularity, may also contribute to an outbreak's trajectory and persistence

(18,19). Additionally, network structure, social support, and sociability may impact individual susceptibility to infection (20,21).

Importantly, individual and community traits can interact with network structure, affecting infectious disease dynamics. Recent measles outbreaks (2013, 2019) in Orthodox Jewish communities in Brooklyn, for example, show how insular and close-knit communities can reinforce community vaccination norms, creating environments conducive to rapid and extensive transmission (22,23). Conversely, increasing HPV vaccination rates in Scotland were recently linked to a "considerable and sustained" decrease in cervical disease (24); in effect reducing potentially infectious sexual contact networks by altering individual risk. Additionally, reactionary behavior change may alter an outbreak's trajectory by changing network structure; systematically encouraging and exploiting these behaviors as non-pharmaceutical interventions may be a practical and key strategy in mitigating outbreaks. For example, reduced movement or travel and isolation after the start of an influenza outbreak have been shown to limit transmission opportunities by reducing network connections (25–28). These results suggest that both network structure *and* individual behavior may be important targets for intervention. However, human factors that can affect transmission patterns remain underexplored (14). There is thus a pressing need for further investigation into the intersections of infectious disease dynamics and social, cultural, and behavioral factors (14). Material gaps persist in attribute and behavioral social network analyses as well as social network data collection and analysis methods—these issues limit the accuracy of current infectious disease predictions and public health's ability to control epidemics. Additional research in these areas will be a crucial step towards developing comprehensive real-world predictions of transmission dynamics.

In addition to affecting transmission dynamics, social networks can impact individual health behaviors that are associated with chronic disease risks. Researchers have explored complex patterns of social network influence and selection on individual alcohol use (29–38), smoking (39–42), drug use (43–45), exercise (46–48), diet and obesity (49–51), among others. Combinations of larger social structures, close relationships, and individual psychosocial and genetic factors may make some individuals more or less susceptible to peer influence (52–55). Conversely, networks may form around shared traits and behaviors, creating an environment that may reinforce behaviors and facilitate their adoption across a contact network. Previous research has identified homophily, which may indicate selection, in many empirical networks, including sexual partnerships (56,57), and adolescent networks (58). Indeed, preferential mixing may be the result of complex interactions along multiple lines of similarity. Block and Grund found that friendships among school-age children patterned according to gender, ethnicity, socioeconomic status, and "pocket money" but that gender and ethnicity exhibited negative additive interaction (59). A deeper understanding of the selection and influence processes involved in social networks will be critical in further disease prevention.

## 1.3 Social Networks and College Alcohol Use

As noted, health behaviors have been extensively addressed in the social network literature. In particular, the associations between peers, contact networks, and risky health behaviors including drug and alcohol use have been investigated, primarily in adult and adolescent populations. Adolescent behavior and network research has identified consistent, positive associations between individual drinking and peer drinking behavior (29,30,32,60,61). However, despite the significant issues with alcohol use and binge drinking on college campuses, social network alcohol studies in this population remain limited.

### 1.3.1 Alcohol use and its consequences

Alcohol use in college students is a major public health issue in the US, as excessive consumption has contributed significantly to morbidity and mortality in young adults (35,62–68), and negative health effects of college drinking can extend well into adulthood (69). College students tend to drink higher quantities and more often than non-college students in the same age group (70–73); approximately 80% of full-time college students report drinking at least once and 40% reporting *regular* drinking (63,74). Of those who do drink, approximately 25% drink at least ten times per month, 50% drink in order to "get drunk," and a third meet the DSM-IV criteria for alcohol abuse (75–78). Moreover, half of all college drinkers are under the legal drinking age of 21 (79). In 2015, 32% of college students reported binge drinking (72) (generally defined as $\geq 5$ (men) or $\geq 4$ (women) drinks on a single occasion (74) and binge drinkers account for between 48 (79) and 91% (80) of all alcohol consumed by college students. Even more alarming, studies have found that over half of binge drinkers are "heavy binge drinkers" ($\geq 7/\geq 6$ drinks (men/women)) (81,82).

The negative consequences of college drinking are extensive and range in severity (77). Students who drink often experience academic consequences, such as missing classes and lower grades (83,84), and short-term health effects (i.e., hangovers, memory loss) (77). Risky behaviors, such as increased numbers of sexual partners, unprotected sex, prescription drug use/abuse, and driving under the influence often accompany drinking (77,85,86) and alcohol has been associated with increased risk of injury and/or being the victim of an assault. Researchers estimate that over close to 600,000 students are injured, 100,000 students are sexually assaulted (with another 100,000 reported being too drunk to know if they consented to sex), and 2.7% of drinkers considered suicide each year (62,70,84,85). Additionally, over three million college students drive under the influence, 1,800 students die in alcohol-related accidents, 700,000 are assaulted by a student who has been drinking, and 5% of college students are involved with police or campus security as a result of drinking each year (77,85). Despite administrative efforts to reduce drinking levels, students tend not to be motivated to curb drinking by past negative consequences or the potential for future consequences, in fact, some "negative consequences" can actually be seen as desirable in this population and may motivate increased alcohol use (87–90).

### 1.3.2 The social context of alcohol use in college

Alcohol use in college students differs from alcohol use in other populations in the US. The changes in living environment, level of independence, and social support systems between high school and college make college students highly susceptible to social influences as they navigate this transition (78,91). Many students begin drinking in their first year of college (45,92–99), when students may be particularly vulnerable to peer influence in order to form social connections quickly (100). Indeed, peer influences have been found to be the strongest predictor of behavior, particularly alcohol and other substance use, in young adults (29,47,91,100–114). The majority of young adults report a social motivation for drinking, including facilitating friendships and social connections, and high peer approval has been associated with heavy drinking (36,112,115–117). As such, heavy drinkers in college tend to do so in groups as opposed to drinking alone to cope with stress (36,118). In the university setting, non-drinkers tend to be viewed negatively (e.g., aloof, antisocial) by other college students (119), providing additional motivation for students to drink in order to be accepted by their peers. The visibility of drinking on campus may also facilitate new friendships according to alcohol use, as students easy identify and form connections with students whose drinking behaviors align with their own (5). As such, membership in organized groups, including fraternities and sororities and athletic teams tend to be associated with higher levels of drinking (94,120–124).

### 1.3.3 Social network studies of alcohol use

For the most part, analyses of peer effects on college student drinking has previously been limited to individual-level predictors (e.g., perceived social support), which use measures of peer behaviors and connections (i.e., "egocentric" network analysis), or context-level predictors, such as residence or fraternity membership. These variables can provide information from the perspective of individuals, but they do not allow identification of network structures that "transcend" individual-level determinants of alcohol use (125). Network data allow for detection of the structural characteristics of individuals and social networks in relation to behavior (126). An individual's type, strength, and number of friendships and their larger social environment may individually and collectively play a role in predicting alcohol use, and conversely, alcohol use may play a role in friends selection (29,100,127). Group-dependent variables cannot be measured without more complete social network data that maps individuals onto a larger network structure.

Social network analyses of alcohol use in non-college populations strongly suggest that social networks structures and relationships are key for understanding alcohol use and changes in drinking behaviors. Specifically, studies in adolescents and adults support a social influence on alcohol use, alcohol's impact on social network formation, as well as the co-evolution of alcohol use and changing social networks. There is a growing body of research on adolescent alcohol use and social networks, mainly utilizing data from the National Longitudinal Study of Adolescent Health (Add Health), a longitudinal, four-wave study of nationally representative adolescents (128). Findings from these studies have repeatedly shown reciprocal effects between network structures (e.g., friendship formation), individual network position (e.g., closeness centrality, degree), and alcohol use (31,38,114,129–134). The consistency of the results across various definitions of friendship, peer group, and friendship strength strengthens the evidence for social influence on alcohol use in adolescents. Interestingly, there seems a measurable influence not only from close friendships but also unreciprocated network links, suggesting that adolescents may drink as a way to strengthen a friendship or gain favor with more "popular" individuals. Indeed, these results suggest that highly connected individuals in a social network may have disproportionate influence on their peers that leads to higher levels of drinking over time (10). While these social network studies provide clear evidence of the bidirectional relationship between social networks and drinking in middle and high school age children, college's unique combination of a highly social environment and ready access to alcohol make it difficult to generalize these findings to college students. In particular, the transition to independent living (or cohabitation in a dormitory), a new educational and social environment, mixing of students below and above legal drinking age, and the ubiquity and availability of alcohol on campus likely to result in different social network patterns of drinking behaviors. For these reason, social network studies within college age populations are strongly needed.

A handful of social network analyses of alcohol use in college students exist, although these networks have been small and the majority were cross-sectional networks, and therefore were unable to distinguish between peer selection and influence. In 2011, Phua's analysis of a dynamic, 3-year network of 34 fraternity members found evidence that alcohol consumption diffused through the network over time and that popularity (indegree) was positively associated with higher daily alcohol consumption (135). In another study, Dumas et al. recruited small "natural drinking groups" of 3-5 same-sex individuals who were leaving bars and surveyed group members individually about within-group relationships and statuses (e.g., ranking group members by likability) and drinking behavior, as well as administered Breathalyzer tests (136–138). Across three publications, they identified a number of positive associations between higher within-group status (e.g., most well-liked by others) and alcohol consumption, as well as higher status and encouraging other group members to drink. Although these analyses utilized a relatively small sample and self-selected groups of drinkers, it highlights the utility of collecting network information from multiple group members in conjunction with individual level data, as well as the significant relationship between social status, peer pressure, and alcohol use. In a study of randomly assigned roommates (dyads, i.e., two individuals per network), Duncan et al. observed that when two male roommates who had been drinkers in high school were assigned to live together, they increased their drinking over the course of the school year (43). Similarly, Eisenberg, Golberstein, and Witlocks' analysis within randomly assigned on-campus roommates found alcohol-peer associations in both men and women, as well as a positive association between pre-existing risky behaviors and the closeness of friendships between roommates (139). In a study of a larger student network (N=125) living in a single college dormitory, associations between number of days of heavy drinking and increased outdegree and betweenness (i.e., the degree to which an individual is a bridge between groups within a social network) were identified and the results suggested peer drinking had a larger influence on behavior than individual-level risk factors (140,141). The use of only residents from a single dormitory limits representativeness of these findings and requires replication in larger study samples. Indeed, research has shown that there are differences in drinking behaviors across various residence types, even within on-campus residence halls. One larger social network study of college students has been conducted. In Belgium, Lorant and Nicaise measured two distinct networks comprised of students in two large classes (psychology and engineering lectures) (142). The found several social network features (in-degree, social capital, network density) were significantly associated with drinking habits and binge drinking—illustrating that network-level features, in addition to individual characteristics, can affect drinking behavior. Unfortunately, reported connections between classmates within large lecture classes may not accurately represent social networks outside of class. Additionally, due to the differences in legal drinking age and habits between Europe and the US, this study's generalizability to US college students is limited.

Overall, there remain major gaps in the current body of research addressing social network effects on alcohol use within college student populations in the US. Despite significant evidence of the major role social context and influences have on alcohol consumption in this population, the majority of studies have examined social influences from the perspectives of individuals and their perceptions of others' behavior. As such, these studies fail to address the larger social context of college student drinking. Similarly, small-scale network studies, such as roommates, single dormitory, or small friend groups, lack the power to examine relationships between individual alcohol use and larger scale network structures, such as density, full network behavior, such as total network alcohol consumption, or individual position within the larger social network structure.

## 1.4 Social Network Structures and Characteristics

Generally, social networks are comprised of edges and nodes, and in combination, they form a 'network' of individual units and connections between them. In order to describe and analyze these networks, we must define a number of structures and characteristics generally, and in terms of their use in this dissertation.

### 1.4.1 Network structures and objects

*Node:* an individual or unique vertex in the network; could represent e.g., a study participant, an organization, a geographic location.

*Edge:* a connection between two nodes in a network; could represent e.g., face-to-face contact, sexual contact, electronic contact (e.g., an email or tweet), a familial connection, a type of relationship (e.g., friend), and so forth.

*Network (sociocentric):* a set of nodes and edges that define a set of relationships or lack-thereof between the set of nodes. A network can also be referred to, especially in the mathematical literature and in reference to network visualizations and characteristics as a "graph." Given the circular dependence within the network (i.e., the network structure relies on observations from multiple (or all) nodes), parametric statistical methods are often invalidated by the non-independence of observations (143).

*Egocentric network*: a "network" with no interconnections between star-shaped clusters of nodes (often referred to as "egos" in this type of network); each individual ego-network consists of a single node and a set of contacts or related individuals ("alters") directed linked to them and only them (i.e., the egos contacts are not surveyed). This type of network assesses "personal" networks as individual units; data is not collected on any edges or connections that would connect egos to each other. As such, this type of network does not necessarily violate the independence assumption of most parametric statistical approaches but cannot be used to assess larger network structures or contact patterns (144).

*Network boundary*: the set of observed nodes in a given measured network.

*Dyad:* a pair of nodes in the network connected by one or more edges.

*Neighbor:* a node directly connected to a given node by one or more edges.

*Simple graph*: a network with a maximum of one edge (or one edge in each direction) per dyad.

*Multigraph*: a network with possible multiple edges per dyad (e.g., collected at different time points, different types of interpersonal relationships).

*Directed network (or digraph):* a network where edges have directionality, in that they have a node of origin and a destination node. I.e., contacts emanate from a node and are pointed, or directed, at another. Some network analyses also allow single edges to be bidirectional.

### 1.4.2 Node characteristics

*Centrality:* a general term for describing the level of connectedness or importance of a node; a number of specific centrality measures (described below) can be used to describe or identify nodes that are important to the network's structure and topology and/or potentially influential in terms of transmission patterns across the network (145,146).

*Degree:* the number of edges connected to a given node (can be defined to include directional edges or only be a count of neighbors); a centrality measure.

*Indegree:* in a directed network, the number of directed edges that are "pointed" at a given node; a centrality measure that is often used as a proxy for "popularity" in a network (135).

*Outdegree:* in a directed network, the number of directed edges that emanate from a given node to that node's neighbors; a centrality measure often interpreted as a measure of extroversion or gregariousness (147).

*Isolate:* An isolate in a network is a node without any ties or edges to other nodes (i.e., a node without any neighbors).

Dyad reciprocity (node level): in a directed network, the proportion of a given node’s neighbors with whom they have reciprocal edges.

*Clustering coefficient (local):* A measure of how many of a given node’s neighbors are connected to each other, or the proportion of edges that exist out of the possible edges between all of a node’s neighbors; an individual measure of local density and transitivity (148).

*Closeness centrality*: A node’s *closeness* refers to the average shortest distance to all other nodes in the network (i.e., the average number of edges between a node and all other nodes); nodes with high closeness have short distances to all other nodes in the network and thus can efficiently transmit information or infections to other nodes (149,150).

### 1.4.3 Edge characteristics

*Reciprocal:* In a directed network, either a single edge that is directed at both members of a dyad or a pair of oppositely directed edges in a dyad; reciprocal edges are often considered stronger or more resilient to disruption than unreciprocal edges.

*Unreciprocal:* In a directed network, an edge that is only directed at one member of a dyad.

*Weight:* A weighted edge indicates the “strength” of an edge according to a defined property, e.g., contact duration, number of different relationships shared by a dyad.

### 1.4.4 Network characteristics

*Density:* Density is a measure of network connectedness, quantified as the ratio of the number of edges in the network to the total possible number of edges in the network (i.e., if every node in the network was directly linked with an edge to every other node in the network).

*Transitivity*: A network’s *transitivity* is a measure of network clustering. It is quantified as the proportion of triangles in a network (i.e., three nodes all connected to each other by edges) out of the number of possible triangles in the network (151).

*Dyad reciprocity (global):* the proportion of all dyads in the network that contain reciprocal edges.

*Clustering coefficient (average):* the mean of the network's nodes' individual clustering coefficients; a measure of how tightly clustered the network is (148).

*Homophily:* a general term to describe the extent to which a network exhibits selection, or preferential contact between similar nodes.

*Assortativity:* The *assortativity* of a network is a homophily statistic that describes the extent to which a network's edges are preferentially attached according to node traits. The assortativity coefficient (*r*) for degree assortativity is the Pearson correlation coefficient between pairs of connected nodes, averaged over the network (17). For attribute assortativity, *r* can be interpreted as similar to an intraclass correlation coefficient (57). Assortativity by a given attribute is calculated based on a matrix of mixing probabilities for the attribute of interest (i.e., the probabilities of all combinations of an attribute for two directly linked individuals). Assortativity values between 0 and 1 indicate preferential attachment of similar individuals and values between -1 and 0 indicate preferential attachment of dissimilar individuals (disassortativity). The assortativity coefficient (*r*) for a given attribute can then be calculated according to: $r = \frac{\sum_i e_{ii} - \sum_i a_i b_i}{1 - \sum_i a_i b_i}$, where *i* represents the categories of an attribute, $e_{ii}$, the diagonal elements of a matrix of mixing probabilities for an attribute, $a_i$, the row sums of the probability matrix, and $b_i$ is the column sums of the probability matrix (57). Assortativity is described in greater detail in 1.4.5.

### 1.4.5 Social network assortativity

Assortativity, or assortative mixing, describes a network's homophily, or the observed propensity for a network to be more densely connected among similar individuals. Assortativity has been deemed a key property of social networks in particular (as opposed to other types of networks, e.g., animal, neural) (152,153), as it can impact a network's systemic behavior and resilience (17). In many social network analyses, assortativity is used as a generic term to mean "degree assortativity," or the average level of correlation between a node's degree (i.e., connections) and that of their neighbors; skewed degree distributions which are often observed in real-word networks are thought to be a result of (degree) assortative mixing (17). If a network has values of assortativity greater than 0 (bounded by 1), this indicates that nodes with high degree will preferentially attach to other nodes with similarly high degree and less socially connected nodes will be more likely to be attached to other nodes with few connections. Intuitively, we can see that networks assortatively mixed by degree will have core clusters of highly connected individuals which are connected to satellites of less well-connected clusters of individuals. These central cores are particularly relevant in terms of transmission and social influence as they will be resilient to change and disruption (e.g., removing a single node will not disrupt the cluster as a whole). In terms of transmission, these dense clusters may sustain transmission in ways the complete network cannot, presenting significant barriers to eradication (17,56,154). Socially, these clusters may have disproportionate influence over network behavior and the spread of behaviors; alternatively disassortative degree mixing has been suggested to increase how quickly an idea or behavior is spread across the network as a whole (152). Research has found that empirical social networks tend to exhibit assortativity (153) and that random networks can exhibit degree assortativity (155). This fact makes further investigations of assortative mixing in real-world networks and the degree to which assortativity above that which could occur by chance particularly important.

Assortativity, while often used to describe the overall degree correlation between dyads, can also describe preferential attachment by scalar (e.g., age), enumerative (e.g., race), and/or behavioral (e.g., vaccine acceptance) attributes. Attribute-based assortativity describes both homophily (the general term for networks clustered according to similar traits) and selection (the deliberate formation of edges according to individual traits). Logically, human interactions pattern along individual characteristics and real-world networks can exhibit assortative and disassortative mixing depending on the network, its measurement, and the attribute in question (57). These patterns may occur by default; for instance, a racially homogenous community's contacts will necessarily fall within these racial constraints. However, even in a racially heterogeneous community, social contacts may exhibit racial segregation if on the whole individuals preferentially connect according to race. Indeed, human interaction tends to pattern along multiple traits and their interactions (59) and preference patterns may differ temporally, as well as according to relationship or interaction type, location or environment, necessity, and/or situation. For example, individuals may prefer one gender for their sexual relationships and another for friendships and these preferences may shift over time and depending on the individual's age, marital status, and so forth. Generally, network-wide patterns of preferential segregation can lead to social stratifications and clustering (57); the sum of individual social preferences can impact holistic network behavior and topology.

## 1.5 Social Network Measurement

In this dissertation, we are concerned primarily with in-person contact networks where nodes represent individual study participants and edges correspond to in-person contacts of some kind (e.g., close proximity, face-to-face communication). Historically, this type of in-person contact data has been collected for use in infection disease epidemiology (9); the use of "contact tracing" is widely known for its use in tracing the transmission of sexually transmitted infections, such as HIV. Increasingly, social contact data is being used in public health to describe how behaviors and ideas can be transmitted across networks of individuals and communities; researchers are adopting traditional infectious disease methods for use in social epidemiology. Here we outline the typical procedures researchers may use to identify and enroll network members in sociocentric studies and the methods by which social contact networks can be measured and observed. Note that these procedures may vary from those used in egocentric network studies.

An *a priori* identified potential network boundary consists of the set of identified individuals that *could be* included in the network. These boundary specification and identification strategies, each with their own strengths and drawbacks, can be defined according to: 1) formally defined positions or membership, i.e., "positional boundary specification" (e.g., students in a middle school, CEOs); 2) subjective and often self-determined inclusion in a group, i.e., "realist and nominalist boundary specification" (e.g., Democrats, feminists); 3) shared participation in an event or activity ("event-based strategies"), which identifies individuals who were present at an activity occurring at specific time(s) and place(s) (e.g., people who went to a given concert; in some ways this is similar to how public health officials identify exposed individuals during an outbreak investigation); and 4) "relational boundary specification" where a group of initially identified individuals select their contacts and the next "wave" of individuals to include in the boundary, e.g., respondent-drive, chain-referral, or snowball sampling (156). The first three methods may be limited depending on the numbers of identifiable connections between individuals and may have small overall sample sizes, depending on the boundary definition. This final, respondent-driven method is most applicable to this dissertation and is common in public health network research as it tends to produce tightly connected networks (156,157). This sampling method is particularly useful for infectious disease transmission and social influence research as it can identify typically hard-to-identify individuals or groups that may not be sampled or approachable by traditional means (e.g., sexual partners, members of insular communities, homeless individuals) (158,159). However, there are drawbacks in terms of privacy and ethical issues (e.g., identifying sexual partners) as it may provide researchers with information about individuals without their prior consent. Additionally, it can bias the sample away from the underlying "true" contact network of a population, as this type of sampling is by definition a nonrandom process (160), although new approached are in development to address this issue statistically in analyses (161,162). As with all public health research, researchers should clearly articulate their enrollment process and inclusion criteria when communicating results and methods to allow for critical review of underlying assumptions, inferences, and generalizability.

Once the *a posteriori* network boundaries are defined (i.e., participants are enrolled), contacts between network members (within the boundary) can be measured in a number of ways, depending on the research aim. The selection of a network measurement method is quite important, as the measurement tool is intrinsically linked to the measured network's structure, relationship to the underlying "true" network, and therefore the inferences drawn from it (163–172). Here we focus on the collection of in-person contacts between network members (as opposed to other types of direct contacts such as electronic contacts or co-authorships). Few studies rely on direct observation (or watching a recording) of in-person contacts (13); generally, this type of social contact data and networks have been measured with surveys or diaries in which individuals report their in-person contacts and/or relationships with other network members. Broadly, contact diaries or surveys involve asking participants to report contacts that occurred during a certain time-frame and fit a given definition of "contact." These contact definitions vary according to research aims and could restrict reports by location, type of contact,

contact duration, and/or relationship. For instance, study participants could be asked to report face-to-face contacts of a certain duration, contacts that involved physical contact, or in-person contacts with family members. In infectious disease transmission network studies, researchers may ask specifically about contacts relevant to a given pathogen and transmission mode, e.g., sexual contact for sexually transmitted infections or close-contact for droplet transmission (13). Social and behavioral studies may instead request information about influential or important contacts.

A number of simple and complex contact survey or diary types are typically used in contact measurement. These diaries or surveys can be administered on paper or electronically, both retrospectively and prospectively (or in “real-time”), and query contacts, as well as any additional details about the contact that may be used in network construction and edge weighting (e.g., contact duration). Contact reports are elicited in a number of systematic ways, each with its own strengths and limitations that effect the quantity and quality of contact data (see Van Hoang et al. for an in-depth review of contact survey designs (171)). Paper diaries versus electronic reporting may also impact the reporting, depending on participant’s comfort level with the specific technology (e.g., mobile phone apps, web-based surveys, paper diaries). Previous studies have found that paper dairies had higher reporting than hand-held devices (169,173); increasing intrusion of technology in daily life may alter this result. Retrospective reporting requests participants recall their contacts across a defined timeframe whereas prospective measurement requests “real-time” reporting of contacts (i.e., participants record a contact when or soon after it occurs in a paper or electronic diary). Retrospective and prospective designs are both common among contact surveys (171), and some longitudinal network studies involve pseudo-combinations of the two wherein participants are asked to recall contacts at multiple time-points (e.g., Aiello et al. (174)). Retrospective designs may suffer from recall issues, particularly in terms of detailed information like duration, and may be more mentally taxing for participants (171) whereas prospective designs require higher commitment and engagement, possibly limiting active participation among individuals who find reporting burdensome. However, prospective designs may result in higher numbers of reported contacts (171). The referent time frame for recall as well as the length of the prospective study may also improve or diminish data quality and quantity. Additionally, if participants are aware that they will be asked about their contacts, they may take proactive and deliberate steps to improve their recall or make behavioral adjustments to accommodate real-time reporting.

The simplest type of retrospective contact surveys ask participants to list their contacts that fit the outlined definition (e.g., Barabasi and Albert (175)). Other surveys ask for contact reports but limit the number of contacts an individual can report ("fixed response") and others provide participants with a list of possible contacts from which to choose (surveys can also be combinations of these methods). We note that each of these methods may bias responses in some way. Fixed response surveys limit reporting, and therefore may have lower network completeness (170) or may lead to extraneous reports of invalid contacts in order to "complete" the survey. Name generator-based surveys may bias reports to those listed by the generator and free reporting may result in under- or over-reporting, depending on the individual. Some studies incentivize complete data collection by suggesting to participants that their contacts (and in some cases, both the participant and contact) will receive a reward when they too provide their contact information (e.g., Branas-Garza et al. (176)). These studies tend to be used for social and behavioral studies, as they maximize reporting of "important" contacts to those participants care about providing a reward (163). However, this may limit reports to contacts that participants believe will also complete the survey. Similarly, if the reward is limited in number, this may also limit the number of reported contacts. Regardless of the survey design, there is evidence that participant and contact attributes can affect the likelihood of reporting a given contact in either direction, depending on the combination of attributes and the characteristics of the contact event (13,163,171,177,178). Importantly, we also note that reporting and naming contacts (i.e., not anonymous reporting) requires knowledge of the contact (particularly names) on the part of the respondent, and in the case of sociocentric studies, may require participants to be aware of who else is participating in the study and data collection. As expected, the most common form of missing network data is missing edges or reported contacts between participants (179,180). However, the possibility of reciprocal reports, i.e., duplicated reports from both individuals involved in a contact event, can provide some data validation and can somewhat balance contact underreporting or nonresponse by some network members (181). Overall, while valuable, participant-reported contacts are prone to observation/measurement error, impractical for large networks (182), underestimate the true connectivity of the networks in almost all cases, and therefore do not provide a complete picture of real-world contact networks (172,181,183).

As researchers have accepted that empirical contact networks measured via contact reporting are most likely incomplete, methods for objectively collecting contact data have been developed. In order to capture more contact data (in terms of volume and detail) and therefore observe a greater portion of the "true" network, researchers are increasingly using wearable sensors (e.g., "motes," "tags," mobile device apps) to capture people's dynamic social networks using radio, Bluetooth, or cellular communication, across a variety of settings (174,184–190). Sensors capture proximity contacts between individuals by recognizing and recording nearby sensors of the same type at defined intervals, thus allowing researchers to collect fine-grained contact data on contacts between individuals wearing or carrying the sensors were in close proximity to each other. Depending on their specific technologies, sensors may be able to collect data with temporal and spatial resolutions from as fine as 20 seconds and as close-range as 1-1.5 meters (186). Thus, sensor data provides an objective measure of human interactions within a specified network boundary (those carrying the sensor) that does not rely on recall or reporting. Additionally, sensors may also incorporate technology to record detailed biometric and/or environmental data (e.g., pedometers, temperature and humidity), and hypothetically may be placed on objects or in geographic locations to record proximity-based interactions with inanimate objects of interest (e.g., utilization of hand soap dispensers) and/or movement patterns within defined areas. However, sensor contact-detection methods have a number of practical challenges (182,191,192), including: 1) capturing large amounts of data that can be difficult to organize and analyze (192); 2) participant privacy concerns (193); and 3) the probability of capturing spurious (e.g., across a wall) contacts (13). Additionally, sensors may capture *too much* data, making it difficult to parse and identify the contacts relevant to transmission. Sensor studies are also practically and financially limited to smaller networks, thus only capturing contacts between a relatively small number of participants wearing active sensors (13). While survey methods capture fewer contacts, they have the advantage of being comparatively easier studies to run, can be run pro- or retrospectively, and can collect egocentric contact data as well as hypothetically complete, within study network data (13). Despite the promise of sensor-based network measurement as a more accurate epidemiological tool than surveys, little is actually known about how sensor data compares to surveys in terms of completeness, validity, and practical relevance to transmission modeling.   As such, comparing networks collected via contact surveys and sensors is a key step in determining if there is an identifiable relationship between contact diary and sensor-data. By investigating the characteristics of, and agreement between, differently observed networks we can speculate as to the practical and intrinsic strengths and weakness of each approach, as well as their comparability.

Before widely adopting the technology, researchers must compare sensor networks and data to the current (flawed but practical) survey methods. Given the early stages of sensor technology utilization in network measurement, it is understandable that little is known about the characteristics of sensor-captured networks, their relationship to underlying "true" networks, and their long-term utility in epidemiological research. As far as we are aware, only a handful of studies have been published that directly compare diary and sensor contact data: Smieszek et al. (2014) directly compared web-based survey and sensor contact data (including contact and duration) collected over the course of three consecutive days in a US high school (187); Leecaster et al. compared diary and sensor networks collected on a single day in a US middle school (185); Mastrandrea, Fournet, and Barrat compared networks collected via survey and sensor on a single day in a French high school (184); and Smieszek et al. (2016) investigated differences between contact diary and wearable sensor data collected during one day of a public health conference (194). Each of these studies was short-term and had a limited network boundary (e.g., participants at the conference).

In all four comparison studies, a number of network metrics were compared between data collection methods, including participant degrees, network density, and the degree distribution. Generally, these studies found that surveys tended to collect fewer contacts, and therefore less dense networks, than sensors. Smieszek et al. (2014) found that contact underreporting accounted for 30 and 50% of the discrepancies between the two networks. However, these discrepancies were reduced when the shortest sensor-contacts were omitted from the network, i.e., reporting accuracy increased with contact duration. Additionally, female participants reported more "true" contacts than male, and higher degree (sensor) individuals' reports more closely matched sensor data than lower degree participants. However, they did not find a correlation between individual degrees measured by sensor and self-report. Their results suggest that the two measurement methods cannot be used interchangeably, but restricting the contact definition to longer durations and removing lower-degree individuals from the network can limit differences between networks. Leecaster et al.'s study in a US middle school (185) also found large discrepancies between the two. Specifically, they found that the sensor network degrees were approximately three times higher than reported network degrees. Mastrandrea, Fournet, and Barrat's comparison of a single day's collection of sensors and reported contacts in a French high school (184) determined that their observed sensor network was approximately twice as dense than the diary, although the two networks had similarly shaped degree distributions. Smieszek et al.'s conference-based comparison also highlighted that neither data collection method resulted in "complete" networks; while there was overlap between reported and sensed contacts, neither was a complete subset of the other. Unlike Smieszek et al. (2014, 2016) and Leecaster et al.'s studies, Mastrandrea et al. concluded that despite differences, the underlying network structure was similar between the two methods. Additionally, Mastrandrea et al. noted that an individual's indegree (contact reported by other participants) more closely resembled their sensor degree than their outdegree, reinforcing importance of collecting sociocentric network data and not just individual (i.e., egocentric) data. They additionally asserted that missing data in diary networks could be imputed using the relationship between the diary and sensor-measurements and the resultant network should approximate true contact networks, which would reduce the need for complicated sensor studies; a recent follow-up paper (195) outlines these methods and Fournet and Barrat used the same data to assess how differences between the two networks impacted the results of transmission models run on the networks (196).

Generally, although these limited studies were consistent in their finding that sensors and diaries/surveys result in topologically different networks, they showed a lack of consistent specific results in terms of overall agreement levels and whether there was an identifiable relationship between differently-captured networks. Additionally, it is difficult to generalize their results to the general population, as primary and secondary schools constitute a "closed" population that is relatively stable across time (within a given year), participants retrospectively reporting contacts may have been able to better remember these contacts, given their stability over time, than those outside of a school setting. Wang et al. found that the vast majority of school-transmission of H1N1 in 2009 occurred within-grade (197) and all three school comparison studies in children noted high assortativity within classes and grades, as well as patterns associated with lunch breaks and between-class breaks, suggesting stable within-school networks. Indeed, this highlights the strong need for a comparison of network data collection methods for post-secondary school networks, as the secondary school networks examined thus far were all highly assortative by school schedule and grade. This makes it possible to construct fairly realistic networks using only school schedules and rosters, which is unlikely to be possible for college or adult populations, where social interactions happen in more varied and non-closed populations. Additionally, it is unlikely that within school contact patterns mirror those after school, for both children and adults. School days are highly structured and tend to be comprised of blocks of time spent in close proximity (i.e., classrooms, hallways) to a fixed set of similar individuals, a pattern that is not typical of other environments. Even barring the context-difference considerations, research has repeatedly found distinct age-related contact patterns (56,198,199). Therefore, associations between the two network types in these studies in adolescents may not hold in other populations and settings. However, these previous sensor/diary comparisons provide a framework on which to base additional research.

Thus, additional research focused on understanding concordance and relationships between networks using contact diaries and sensors in other populations—particularly longitudinally—is warranted. For instance, there are no published investigations comparing long term patterns in empirical network data collection between sensor and diaries; there is a substantial gap in the dynamic networks literature on the benefits and drawbacks of these different instruments. Given the importance of social network analyses in infectious disease and behavioral health research, there is a clear, continuing need for comparison studies in order to further clarify the strengths and weaknesses of each individual data collection method, and relationships between their data.

### 1.6 Uncertainty and Hypothesis Testing in Social Network Analysis

As a general concept, uncertainty describes inadequate information related to a lack of knowledge or inherent variability (or both) (200). Given that we cannot compare theoretical or empirical networks to a “true” network, the validity of any network data is intrinsically questionable (172). Thus, we must delve into the potential sources and amount of any and all unintentional or intentional data errors (201,202), conservatively assume that error is a non-trivial issue in network data (201,202), and explore how data errors may cascade or propagate throughout network analyses (203,204). Explorations of network data accuracy and its effects on network statistics (e.g., centrality), have begun but are still underdeveloped overall (205).

Social network data error(s) can come from a number of sources and real world network datasets likely contain a mix of errors (172). Network sampling and observation may result in edge or node addition or deletion errors (160,206), in addition more nuanced errors in detailed social contact data (e.g., incorrectly recorded/reported contact duration or relationship type). As with all data collection, these errors may be intentionally or unintentionally incorporated into network measurements, but in either case they are likely systematic as well as sporadic. Indeed it may be impossible to identify either systemic or sporadic errors as well as any underlying mechanisms for systemic error (206). A growing, but underdeveloped (205) area of network research investigates the robustness of network statistics by modeling the effects of defined types and amounts of measurement error on theoretical "true" network structure, properties, and behavior. In these studies, researchers have hypothesized probable sources of network data errors and mainly used direct comparisons of different theoretical networks (typically generated with permutation and resampling methods) to estimate of how robust and reliable network structure and inferences may be (179–181,202,206–209). This research has shown that different types of network data error appear to have differing impacts on network statistics and characteristics, depending on the error type (206), the statistic or characteristic of interest (206), and the "true" network topology (172). It has also been noted that it may be impossible to separate error's sources and define its effects in empirical as opposed to theoretical networks (209). Additionally, Franz et al. (2009) noted that generally, social network analysis is currently limited by an inability to determine valid confidence limits for network statistics (172). Thus, quantifying the reliability of attribute assortativity will be an important extension of these previous robustness studies.

Social network data generally consists of attribute data and relational data. Attribute data describes node network characteristics (e.g., degree) and is intrinsically linked to the network structure as a whole, and individual demographic, trait, and behavioral data (e.g., age, gender, alcohol consumption). Relational data describes the network structure and topology overall, as well as within dyads and subgroups within the network (e.g., density, edge weights). The combination of these two types of data into a network produces interdependence between individuals, their traits, and the network. Parametric statistical methods, including many procedures for estimating uncertainty and hypothesis testing, are invalid for network data (for the most part), as they violate these method's underlying assumption of independence observations (210). Therefore, development of nonparametric methods is a key ongoing area of research in analyzing social network data.

## 1.7 Dissertation

This dissertation applies and develops social network data collection and analysis methodology. The overall aims of this dissertation were to improve our understanding of: 1) how social structures and contact patterns shape alcohol consumption and use in undergraduate students; 2) drawbacks and data quality issues related to different methods of measuring social contact networks; and 3) how to articulate and quantify network uncertainty and hypothesis testing for trait assortativity.

We examined a number of social network analysis methods and approaches using the social networks captured by the eX-FLU study—a large, two-year social network study of university students. In so doing, we defined and developed a number of novel methods for social network analysis in public health research including a direct network comparison statistic (adapted Kappa coefficient), multiple sensitivity analyses and uncertainty quantification methods, and null models for hypothesis testing of trait assortativity.

### 1.7.1 Aim 1–Social network and alcohol consumption: structural and individual associations

*Project goal: To identify college student contact patterns and structures associated with alcohol use, particularly those that could 1) encourage initiation and continuation of alcohol consumption and, 2) act as barriers to intervention and behavior change.*

Excessive alcohol consumption significantly contributes to morbidity and mortality in US college students. Peer relationships and influence have been repeatedly associated with alcohol use in this population; social network analysis can further our understanding of these relationships in order to design more effective interventions. Using health behavior and weekly reported contact data collected from approximately 800 eX-FLU participants (2 years), we identified numerous consistent relationships between social network characteristics and structures and alcohol use. These results, including the first explorations of the association of longitudinal tie strength and drinking, indicate that alcohol use and social position are significantly associated among college students. Notably, we identified that network position, network alcohol exposures, and relationship strength were associated not only with individual alcohol use but with consumption volume. This suggests a complex relationship between drinking, larger social structures, and proximity to others who drink. Across multiple social connectivity measurements, drinkers were consistently more social and more popular than non-drinkers. While there were consistent results comparing drinkers and abstainers, there were also interesting differences in network position and contact patterns between moderate and binge drinkers, with binge drinkers being less popular and having fewer connections than moderate drinkers. We also examined how students' face-to-face interactions differ according to their contacts' alcohol use, and found that drinkers had more contact with other drinkers than non-drinkers and that these contacts with other drinkers were stronger and more resilient over time than their non-drinker contacts. These results extend and build upon previous social network analyses of substance use in college students by uncovering significant differences in network properties of moderate versus binge drinkers.

Overall, our results build on a growing body of evidence of significant relationships between network structure, social position, and alcohol consumption. There appear to be complex interactions between social standing, demographics, individual motivations, and alcohol use; these associations, as well as their directionality, warrant further study. Our results show the potential impacts that social position can have on individual behavior and vice versa. Notably, we identified numerous network structures and properties that present potentially significant barriers to on-campus alcohol-abstinence interventions. Consumption-related clustering and popularity likely create social environments that encourage drinking initiation and continuation. Hypothetically, interventions directed towards heavy drinking could exploit social embeddedness and popularity differences between moderate and binge drinkers.

### 1.7.2 Aim 2–Social network measurement methods: comparing sensor and self-reported contact data

*Project goal: To analytically compare two social network measurement methods with varying levels of granularity in order understand the unique utility of self-report vs. sensor contact data, using the eX-FLU Study and its sensor sub-study, iEpi, as a case study.*

Little is currently known about the reliability of self-reported contact data and the stability of an individual's contact network longitudinally. The high potential for under- and mis-reported contacts in social network data may have significant impacts on the efficacy of network-based interventions and may bias disease transmission models. As such, comparing static networks collected via contact surveys and sensors is a key step in determining if there is an identifiable relationship between contact diary and sensor-data. Additional comparisons of contact patterns and network structures across longitudinal (dynamic) social networks will also provide information on contact-pattern stability over time and how frequently contact data should be collected. The long-term goal of this line of research is to identify appropriate methods for contact data collection and possible data transformation for behavioral and infectious disease social network studies.

In this project, we compared eight weeks of contact data and their resultant social networks from two eX-FLU sources: weekly self-report and records of Bluetooth proximity captured by the novel sensor app, iEpi. Networks were compared between and within the two collection methods, using overall network structure, dyad, and node characteristics. As no consensus exists as to how to calculate the agreement between networks drawn from the same underlying "true" network of individuals, we developed an adapted Cohen's Kappa statistic for this project. Comparisons between the two collection methods showed that although the iEpi sensor collected a denser network with higher numbers of contacts between study participants, the percentage of reported contacts that iEpi failed to capture increased over the study period. However, despite this result, the overall network agreement increased over the course of the study, although the highest agreement level, occurring in final study week, would still be considered "slight" (as defined by Cohen). When the iEpi network was restricted to contacts with durations greater than an hour, agreement increased into the "fair" range. Importantly, the size and densities of the iEpi networks dropped off rapidly over the study period, suggesting decreasing participant engagement and tolerance of the device over time, whereas the reported network maintained its size and density throughout the study period. In aggregate, these results point to neither network collection method as a “gold standard.” The increased agreement when looking only at longer contacts shows that the two methods may operate under different definitions of contact. Recalling prior contacts appears to depend on duration and an unknown, subjective level of individual significance, whereas iEpi’s program defined contacts strictly on proximity.

Generally, these results suggest 1) a clear and relevant *a priori* “contact” definition should be driven by the research question and project, 2) researchers should not assume that either sensor or self-reported social network data collection methods will generate a complete social network on which to base simulations or interventions, especially given that 3) sensor-based data’s completeness may decrease over time. In light of these findings, researchers should carefully consider which collection method, or combination of methods, could provide them with the highest quality data needed to answer their research questions.

### 1.7.3 Aim 3–Social network assortativity: sensitivity analyses, uncertainty quantification, and hypothesis testing

*Project goal 1: To develop sensitivity analysis methods to understand assortativity bias due to potential measurement error in participant reported characteristics, specifically nonresponse and misclassification.*

*Project goal 2: To define and investigate methods for assortativity uncertainty quantification, i.e., to establish reasonable confidence limits for attribute-based assortative mixing.*

*Project goal 3: To articulate, define, and implement appropriate null models for hypothesis testing in social network analysis, specifically for attribute-based assortativity.*

As the inherent dependence of social network data violates independence assumptions required by parametric statistical tests, statistical uncertainty and hypothesis testing is often overlooked or unreported in health and behavioral social network literature. We investigated methods to address uncertainty in assortativity, a social network measure of clustering, resulting from self-reported social network data, using a variety of randomization methods based on the most appropriate null model for our research questions. We: 1) identified potential sources of assortativity uncertainty, particularly missing or inaccurate reported contact data and implemented resampling methods to incorporate these potential measurement errors into quantifications of assortativity confidence limits; 2) clearly defined two types of null models to control for hypothesized attribute-based assortative mixing and implemented randomization procedures testing the component effects of selection on mixing patterns, determining the statistical significance of multiple attribute assortativities.

We assessed three attribute assortativities of the measured eX-FLU baseline social network (alcohol use, ethnicity, and gender), and developed and implemented multiple approaches for hypothesis testing and analyzing the potential impacts of imperfect individual and network data. Specifically, the sensitivity analyses and uncertainty quantification systematically explored the robustness of attribute assortativity in the face of multiple types of measurement error. For the first time (that we are aware of), we addressed how measurement error and nonresponse of self-reported covariates may bias attribute assortativity. We then used similar approaches to quantify the range of values in which the network's "true" attribute assortativity likely lies, given a range of levels of random network node and edge data error. Overall, we found that attribute assortativity biases and uncertainty may be predictable, given an *a priori* defined amount and type of data error. These methods provide a framework for attribute assortativity uncertainty analyses, as well as hypothesis testing, on sociocentric networks, particularly in contact network studies.

## 1.8 The eX-FLU Study

Data for this dissertation were collected as a part of the eX-FLU study, a social network study conducted at the University of Michigan in 2012-2013. The purpose of the study was to examine transmission of influenza in social networks of students living on-campus, but a large amount of data on other health behaviors, including baseline alcohol use, were also collected. A detailed description of the eX-FLU methods can be found in Aiello et al. 2016 (174) and portions of the surveys relevant to this dissertation can be found in the Appendix. In short, eligible students (≥18 years old) from six coed residence halls at a large, public university were enrolled via respondent-driven sampling, i.e., chain-referral, in the fall and followed for eight ("Pilot" or

“Year 1” study) and ten (“Main” or “Year 2” study), weeks in the following winter semesters (2012, 2013). Study enrollment was done via snowball sampling; initially, students were recruited to join eX-FLU by study staff, and upon enrollment were given the opportunity to invite their friends and social contacts to join the study as well. This process continued for subsequent waves of enrollment, resulting in a large social network of nominated students and enrolled participants. In addition, during the second year of the study, a sample of participants was recruited to also participate in a network-sensor sub-study known as iEpi. These participants agreed to carry an android smartphone loaded with the iEpi app, which acted as a social contact sensor and recorded proximity-based contacts between iEpi participants over the study period. This dissertation will take advantage of some of this study’s unique data, in particular the longitudinal (Pilot: 8 weeks, Main: 10 weeks) self-reported social network data and, for iEpi participants, a dynamic set of objectively collected contacts, captured via Bluetooth-sensor. A detailed description of the iEpi contact-data collection methods can be found in Chapter 3.2.2.

Over the course of the enrollment and study periods, participants responded to multiple surveys. During enrollment and prior to (Year 2) or at the start of (Year 1) the study period participants reported baseline demographics, health status, and health behaviors. During the study periods, participants were emailed weekly surveys on which they were asked to report other study participants with whom they had face-to-face contact with over the previous seven days, which was used to create multiple social contact networks. On these surveys, participants could select social contacts from a list of likely contacts pre-populated with names of previously reported contacts and/or participants could search for other participant-contacts by name or email. No limit was placed on the number of contacts participants could report on each survey.

# Chapter 2

# Social networks and alcohol consumption in college students

## 2.1 Introduction

College alcohol use is a major public health issue, as excessive consumption significantly contributes to young adults' morbidity and mortality (67,68,70,85), the effects of which can extend into adulthood (69,211). Drinking is entrenched in the typical college experience (212); in 2015, 79% of college students had consumed alcohol in the past year and 2-week prevalence of binge drinking was 32% (213). Despite university interventions, overall levels of alcohol use in this population has not declined over the past 30 years (63), suggesting a critical gap in our understanding of the individual, social, and environmental factors that influence alcohol consumption among college students.

Many college students begin drinking freshman year (45,92–99), when their inexperience with alcohol increases the risk of heavy drinking and its consequences, including academic issues and negative health effects (77,211,214). During this key transition period defined by newfound independence, students are particularly vulnerable to peer influences as they navigate new academic and social environments (100). Accordingly, research consistently shows that college drinking is usually socially motivated (91,112,215). In fact, peer influences appear to be the strongest predictor of behavior, particularly substance use, in young adults (84,91,97,101,102,105,107,114,116,121). Alcohol's ubiquity and availability on-campus, in combination with high peer pressure susceptibility, encourages students to develop risky drinking patterns (45,79,99,100,216). Students use alcohol to facilitate or strengthen friendships, initiate sexual relationships, and to overcome social anxiety and low social capital (36,112,115–117,217). There are likely dynamic, bidirectional relationships between social factors and alcohol, as social networks form around similar alcohol use and in turn, students' consumption is influenced by their social network.

Despite established relationships between peer influence, social motivations, and individual drinking in college, social network structures and characteristics associated with drinking in many college settings have not been well characterized. The majority of previous network studies of college alcohol consumption have relied on egocentric network analyses, where analyses utilize individuals' immediate contacts and perceptions of others' behavior without incorporating reciprocally identified relationships or larger social structures. The advantage of sociocentric network studies (i.e., where individual and connection data is collected from all network members within a specified boundary) is the ability to analyze individual, group, and network-level behavior associations. Although there is a growing body of sociocentric college network-alcohol research, previous college studies have been limited by cross-sectional design, small network size, and/or narrowly defined settings and contact-pools (e.g., roommates, a single course/dorm, fraternities or sororities only). There are still major gaps in our understanding of how network position, relationship strength and reciprocity, and alcohol exposure within relate to alcohol use and consumption volume in college populations.

This study adds to the existing body of literature by examining the face-to-face social network characteristics of baseline moderate, binge, and non-drinkers' within two large, weighted college social networks. Within this secondary data analysis, we sought to uncover the most salient contact network characteristics as potential points of future intervention by analyzing associations between baseline alcohol use and extensive social network structures and characteristics, including network density, transitivity, numbers of contacts, popularity, longitudinal relationship strength, reciprocity, and network alcohol exposures.

## 2.2 Methods

### 2.2.1 Data source

Data for these secondary analyses were collected as a part of the eX-FLU study, which was conducted at a large, public Midwestern university during the 2011-2012 (Year 1) and 2012-2013 (Year 2) academic years. In addition to investigating infectious disease transmission in longitudinal student social networks, the study also collected alcohol consumption data to assess whether drinking influenced infection susceptibility. A detailed description of eX-FLU chain-referral enrollment and study methods can be found in Aiello et al. 2016 (174). Briefly, 2 student cohorts (≥18 years), living on-campus in 6 representative residence halls were enrolled via chain-referral in the fall and followed for 8 weeks in Year 1 (N=584) and 10 weeks in Year 2 (N=590) in the winter semester. While distinct study populations were enrolled each year, re-enrollment for Year 1 participants who were still eligible in Year 2 was allowed and 79 students participated both years. We used self-reported enrollment demographics, baseline health behavior (reported at the start of the study periods), and weekly within-study contact reports to conduct a secondary data analysis.

### 2.2.2 Measures

*Demographics and alcohol use*

At enrollment, participants reported gender (male/female), academic year (freshman/sophomore/junior/senior/senior+), race, and date of birth. Date of birth was used to calculate age, which was used continuously and categorically (<21, 21+ years). At the study period's start, baseline drinking was assessed. Current alcohol consumption was measured with a modified version of the Daily Drinking Questionnaire (51). Participants reported if they currently drank alcohol at least once a week and were dichotomized as baseline drinkers or non-drinkers. Drinkers additionally reported how many drinks they "usually consume on each day of an average week (0, 1, 2, 3, 4, 5, 6+; 1 drink equal to a glass of wine, 12 oz. beer, or shot of hard liquor)". Drinkers who reported consuming 5+ (men) or 4+ (women) drinks on at least 1 day were categorized as baseline binge drinkers (74). In Year 1, 16.6% of participants and in Year 2, 29.3% of participants did not provide alcohol use data and were excluded from demographic and nodal analyses. Due to 79 participants' repeat enrollment and different study period lengths, Years 1 and 2 were analyzed separately. Survey excerpts containing the alcohol consumption instrument can be found in Appendix A.3.

*Sociocentric network construction*

Weekly contacts reported during the 8 (Year 1) and 10 (Year 2) week study periods were used to generate social networks containing all contacts within the study population boundary.

Participants were sent weekly contact surveys on which they reported all other participants with whom they had face-to-face contact during the previous week (see contact survey in Appendix A.1). Participants selected an unlimited number of contacts from a list of likely contacts (pre-populated with previously reported contacts) and/or by searching for study participants by name/email (data was de-identified prior to analysis). In order to assess associations between baseline alcohol use and aggregate contact patterns across the study periods, social networks containing all contacts across the study period was created and these networks were analyzed cross-sectionally. Reported contacts were translated into directed edges, emanating from the reporting participant ("node"), directed at their reported contact ("neighbor"); a maximum of 8 (Year 1) and 10 (Year 2) edges could tie each "dyad," or pair or participants. Edges were considered reciprocal if they were reported by both participants in a dyad on a given week's survey. Dyad contact strength analyses utilized weighted, undirected versions of the networks, where all contacts within dyad were collapsed into single edges, weighted by the number of weeks on which they were reported (by either dyad member); this measure represents the frequency of reported contact over the study period. Social networks were constructed using all available data, including contacts with participants without alcohol data. Networks were constructed and analyzed with Python's (2.7) NetworkX package (218) and visualized in Visone 2.17 (Konstanz, Germany).

Multiple social network properties and nodal characteristics with hypothesized alcohol associations were calculated (see Table 2.1 for definitions). For participants, degree and centrality measures were assessed: numbers of neighbors, outdegree, indegree ("popularity"), dyad reciprocity (node), closeness centrality, and clustering coefficient. We also calculated two measures of social exposure to alcohol use: "network exposure to drinking/binge drinking," defined as the proportions of an individual's neighbors that were drinkers/binge drinkers. In order to investigate if network topology varied by consumption level, the full networks were divided into 4 sub-networks containing only participants within each drinking category (all, moderate, binge, non-drinker). Transitivity and density were calculated for sub-networks.

### 2.2.3 Analysis

*Individual characteristics*

Distributions of demographic and nodal characteristics were calculated for participants who provided alcohol use data. Differences in demographic distributions were assessed with Pearson's Chi-squared or Fisher's exact test. In order to account for social network data's inherent interdependence, hypothesis testing for differences in means of social network characteristic (e.g., mean neighbors between drinkers and non-drinkers) were tested using bootstrapped resampling with replacement (219–221). Under the null hypothesis of no difference in group means, the pooled distribution of a given characteristic was shuffled and observations were resampled across groups. For each of the 10,000 resamples performed, the difference in group means was assessed; P-values represent the proportion of resampled differences in means that were as or more extreme than the data's observed difference in means. Distributions were compared between drinkers and non-drinkers, and moderate and binge drinkers.

*Social networks*

In order to investigate if contact strength, as measured by edge frequency and dyad reciprocity, varied across dyads with different drinking behavior combinations, we assessed the mean multi-edges (i.e., "edge weight," or frequency, the number of weeks contact was reported by one or both participants) and mean dyad reciprocity for dyads containing combinations of non-, all, moderate, and binge drinkers. Means were compared across dyad drinking combinations using the bootstrapping method described above.

## 2.3 Results

### 2.3.1 Demographics and alcohol use

In Year 1, 481 participants (82.4% enrollment) and in Year 2, 423 participants (71.7% enrollment) provided alcohol use information. There were no significant demographic differences between participants with and without alcohol data in Year 1, but in Year 2, participants without alcohol data were slightly younger and more likely to be male than those with alcohol data (results not shown). Baseline demographics and alcohol use distributions were similar across study years (Tables 2.2-2.3). The majority of participants in the sample were underage, female, freshman, and White. Each year, approximately 35% of participants were baseline drinkers. Drinkers were significantly more likely to be White and of legal drinking age than non-drinkers. Approximately half of drinkers were binge drinkers (Year 1: 56.3%; Year 2: 50.3%; Tables 2.4-2.5). Significantly higher proportions of binge drinkers were underage and freshman than moderate drinkers. Drinkers reported heaviest consumption on Friday and Saturday, when approximately 50% reported having 4+ drinks (Figure 2.1).

### 2.3.2 Social networks

Network visualizations highlighting drinkers are shown in Figure 2.2. Drinkers were more central than non-drinkers; drinkers had higher mean neighbors, indegree, outdegree, and clustering coefficient than non-drinkers (indegree significant in Year 1; all significant in Year 2). Drinkers also had significantly different alcohol exposures than non-drinkers, indicating clustering by baseline alcohol use. In Year 1, 40% of drinkers' neighbors were also drinkers and 27% were binge drinkers, compared to 31% and 16% in non-drinkers, respectively. In Year 2, 44% of drinkers' contacts were drinkers, and 20% were binge drinkers, compared to 26% and 11% in non-drinkers. Contact patterns also varied by consumption volume. Moderate drinkers had higher neighbors and in- and out-degree than binge drinkers, although clustering coefficients, dyad reciprocity (node), closeness, and alcohol exposures were similar.

Social network structure and properties for the full networks are shown in Table 2.6. Overall network structure, including dyad reciprocity (global), density, and transitivity was consistent across years. Both networks had low density and high transitivity, which together indicate a highly clustered network. Differences in transitivity and density across the alcohol consumption sub-networks are shown in Table 2.7. Drinker networks had higher densities and transitivities than non-drinker networks. Additionally, there were higher transitivities and densities within the sub-networks of moderate drinkers than binge drinkers, indicating higher interconnectedness between moderate drinkers than binge drinkers. In Year 1, the global dyad reciprocity was 0.31 (i.e., 31% of all edges were reciprocal and therefore reported by both parties in dyad on a given week) and in Year 2, the global dyad reciprocity was 0.28. The mean dyad reciprocity for nodes in Year 1 was 0.22 (SD: 0.32) and in Year 2, it was 0.19 (SD: 0.29). Mean multi-edges (i.e., edge weight) for all dyads, regardless of alcohol use, was 4.27 (SD: 2.74; Year 1) and 5.34 (SD: 3.36;

Year 2). In other words, across all pairs of participants who had contact during each year's study period, contacts (i.e., dyads) were reported by at least one member of the dyad on approximately 50% of each year's weekly surveys. In both years, average edge weight and reciprocity were higher within dyads of drinkers than dyads containing at least one non-drinker (significant in Year 2; Figure 2.3). These results suggest that drinkers were slightly more likely to report contact with each other than contact with non-drinkers. When comparing edge weights across possible consumption level-dyads, moderately drinking dyads had the highest contact weight and reciprocity (Figure 2.4). Contacts between binge drinkers and non-drinkers were had lower weights than contacts between moderate and non-drinkers, but had higher mean weights than contacts between non-drinkers in both years.

## 2.4 Discussion

Across 2 large college student social networks, we identified numerous, consistent relationships between social network characteristics and structures and alcohol use. These results, including the first sociocentric network explorations of the association of longitudinal tie strength and drinking, indicate that alcohol use and social position are significantly associated among college students. Notably, we identified that network position, network alcohol exposures, and relationship strength were associated not only with individual alcohol use but with consumption volume. This suggests a complex relationship between drinking, larger social structures, and proximity to others who drink. Across multiple social connectivity measurements, drinkers were consistently more social and more popular than non-drinkers. While there were consistent results comparing drinkers and abstainers, there were interesting differences in network position and contact patterns between moderate and binge drinkers, with binge drinkers being *less popular* and having *fewer connections* than moderate drinkers. We also examined how students' face-to-face interactions differ according to their contacts' alcohol use, and found that drinkers had more contact with other drinkers than non-drinkers and that these contacts with other drinkers were stronger and more resilient over time than their non-drinker contacts. These results extend and build upon previous social network analyses of substance use in college students by uncovering significant differences in network properties of casual versus binge drinkers.

Despite being a minority in the study populations, drinkers were significantly more socially connected, central, and popular than non-drinkers. Given that peer effects consistently appear to be most important predictor of college alcohol use, these results are not highly surprising. However, as our alcohol data was collected at baseline, we cannot identify causal directionality between drinking and network centrality. There may be a self-reinforcing relationship between social factors and drinking, wherein drinkers socialize and drink together, simultaneously perpetuating behaviors and providing opportunities to make new relationships and strengthen existing friendships. Importantly, drinkers' higher indegrees, relative to non-drinkers, indicate that they were not just more social, but more "popular" (i.e., named by more people on the surveys) than non-drinkers. These differences are particularly striking in Year 2, where non-drinkers had approximately 40% fewer contacts over the study period and approximately 25% fewer neighbors than drinkers. Previous research has found that central network members tend to be influential (222), both by directly pressuring others into conformity or by modeling behavior (91); our results show that drinkers were these influential network members. These effects could be present regardless of personal ties to central drinkers, as a network's highly visible, "key individuals" can be as or more influential on individual alcohol use than the network as a whole (223). We would expect that these central, popular drinkers would act as role models in the college community, which may have led to baseline non-drinkers emulating their behavior over the study period, potentially raising overall levels of alcohol use over the semester.

In addition to being popular, drinkers' contact patterns also appear to have been central in structures of the social networks, as we observed connections that were potentially highly conducive to spreading alcohol use to baseline non-drinkers as well as potentially reinforcing continued drinking in baseline drinkers. In both years, the full social networks had low densities and high transitivities, indicating high clustering and cohesion. Contacts also had high mean frequencies and ~30% reciprocity, indicating that contacts were generally strong and stable across the study periods. Generally, these characteristics make the network highly conducive to behavior propagation (46). Importantly, clustering patterned along baseline alcohol use; drinkers tended to aggregate in central, tightly-knit groups, resulting in strong network cores of drinkers. Drinker sub-networks had higher density and transitivity than non-drinker sub-networks, and drinker dyads had higher frequencies and reciprocity than non-drinker dyads. These dense local clusters could be an important part of behavioral maintenance, as cohesive groups strongly enforce group norms and behavior over time; similarity in drinking may help maintain existing friendships (224). These cores of drinkers could also diffuse their behavior to non-drinkers. While non-drinkers had lower direct exposure to drinking, they were not isolated from drinkers. Approximately 30% of non-drinkers' neighbors were drinkers, and these contacts were stronger than their contacts with other non-drinkers. These multiple, repeated contacts with drinkers may have provided an avenue for direct influence on non-drinkers, as individuals tend to be more susceptible to influence when it comes from multiple immediate contacts (225). This influence may have been particularly strong when coming from central, popular contacts. Despite high exposure to other non-drinkers, relationships with drinkers may have primed non-drinkers for drinking initiation over the study period. Therefore, there was likely some conversion to drinking over the study period, particularly among non-drinkers with higher exposure to drinkers. While we were unable to test this, prior research has suggested that non-drinkers often succumb to social pressure and initiate drinking college and that many students increase drinking towards the end of the academic year (226).

In addition to contact pattern differences between drinkers and non-drinkers, centrality varied across alcohol consumption levels. We found that binge drinkers were less connected and central than moderate drinkers. This could reflect that heavy drinking is less socially acceptable than moderate drinking; research in college and adolescent populations has found that heavier drinkers were less appealing and popular than moderate drinkers (222,227,228) and that lower social capital combined with high centrality is associated with higher consumption (142). This could be reflected in the relative differences between in- and out-degree between moderate and binge drinkers. Whereas moderate drinkers' outdegree was 34% (Year 1) and 50% (Year 2) higher than binge drinkers, their difference in indegree, or popularity, was higher, as moderate drinkers had 44% (Year 1) and 71% (Year 2) higher indegrees. In other words, the difference in connectedness between binge and moderate drinkers was heavily influenced by differences in popularity. Despite centrality differences, binge and moderate drinkers had very similar network alcohol exposures. This suggests that there may be individual factors that impact both contact patterns and alcohol consumption volume. These may be related to individual psychosocial and social capital factors, such as impulsivity, anxiety, and/or low self-esteem, which could enhance vulnerability to peer influence and social norms (49,52,53,58). In particular, multiple studies have found that social anxiety and shyness are associated with heavier drinking (229–233). It may be that our networks' heaviest drinkers were using alcohol to overcome a lower ability to socialize naturally but were unable to achieve the levels of popularity held by naturally gregarious students, possibly due to socially detrimental behavior whilst binge drinking.

The results of our research identify important targets and barriers to alcohol interventions for college students. The highly central, visible nature of drinkers' contacts, as well as the popularity of drinkers will likely be a significant barrier to abstinence interventions. Not only are drinkers popular, and therefore presumably aspirational, which influences close and extended contacts, but they might fear losing social standing if they abstained from drinking. In addition, the long-term behavioral reinforcement from their contacts would be difficult to combat. However, interventions focused on reducing binge and heavy drinking might be more effective. Programs highlighting heavy drinking's negative social consequences, including lower popularity, could reduce consumption volume. Additionally, social skills training for socially anxious students could reduce their need to use alcohol to make and maintain friendships.

This study has several limitations and strengths. Generally, our population had fewer drinkers and binge drinkers than previous studies, which have found that between 60%-80% of college students drink at some point during the year (213), and of those, approximately 66% binge drink (68,234). However, our data was consistent across study years and previous research has established the validity of self-reported drinking in this population (63,235). We suspect that the January baseline survey administration may have resulted in our population's low alcohol use (226,236–239). Low binge drinking may have been a product of participants' consumption volume underestimation (240,241). Drinkers underestimating their typical number of drinks could have resulted in binge drinker misclassification and underestimated levels of binge drinking. However, any resultant biases' magnitude and direction would depend on the distribution of network characteristics and any differential underestimation *within* moderate drinkers.

Adolescent studies have shown that influence and selection shape substance use and social networks (242,243). However, with only baseline alcohol use, we cannot say whether our results are due to influence and/or selection. Although we could not examine potential co-evolution and feedback between alcohol use and networks, we did find consistent associations between baseline alcohol use and contact patterns in the collapsed longitudinal networks. Future studies that collect concurrent, longitudinal drinking and contact data will be important in identifying directionality and feedbacks between drinking, individual contacts, and social networks. This is a crucial area for future research as it has been rare for studies to collect both large, dynamic network data and variation in alcohol use over the study period.

Previous research has noted the importance of sociocentric social network studies in understanding the social factors involved in substance use in adolescents and young adults (244). Mapping full network structure allows for identification of how network position and larger structures, in addition to close contacts, impact behavior, particularly in large networks, which can provide more structurally meaningful information than smaller networks (245). While there was undoubtedly underreporting of within-study contacts by participants (185,187), given that self-reported contacts tend be longer and more frequent than those unreported (184), our network of reported contacts likely represent participants' more important and influential relationships within the study population. And although the study's network boundaries did not include the full campus population, network structure was similar across study years, suggesting that college networks of influential/important relationships may be structurally consistent across different cohorts of on-campus students with similar age and school year distributions. These networks, and their longitudinal collection methods, could then be used for future network models and as a template for network measurement.

Overall, our results' build on a growing body of evidence of significant relationships between network structure, social position, and alcohol consumption. There appear to be complex interactions between social standing, demographics, individual motivations, and alcohol use; these associations, as well as their directionality, warrant further study. Our results show the potential impacts that social position can have on individual behavior and vice versa. Notably, we identified numerous network structures and properties that present potentially significant barriers to on-campus alcohol-abstinence interventions. Consumption-related clustering and popularity likely create social environments that encourage drinking initiation and continuation. However, interventions directed towards heavy drinking could exploit social embeddedness and popularity differences between moderate and binge drinkers.

## 2.5 Tables

**Table 2.1** Definitions and Illustrations of Network Components, Structures, and Characteristics Used in Alcohol Analyses.

| | **Definition** |
|---|---|
| Network structures | |
| Node | Study participant, represented by a dot or circle in figures. |
| Edge | A contact between a pair of study participants, reported by 1 or both participants on a weekly survey. If contact was reported on multiple weeks, these edges are collectively known as "multi-edges." A maximum of 8 (Pilot) or 10 (Main) multi-edges were drawn between participant-pairs, corresponding to the number of weekly contact surveys. Edges are represented by lines in network figures, with an arrow directed from the participant who reported the contact to the participant with whom they reported contact. |
| Reciprocal edge | An edge that was reported by both members of a contact-pair on a given weekly survey. Reciprocal edges are identified by arrows pointing at both participants. |
| Isolate | A node (who did not have any reported contact with other study participants during the study period. |
| Dyad | Two nodes connected by at least 1 edge, i.e., a contact-pair. |
| Network properties | |
| Density | The proportion of possible edges within the network that were reported by study participants. The number of possible edges was calculated by multiplying the maximum number of edges in a given week by the length of each year's study period. |
| Transitivity | A measure of how many of a node's neighbors are connected to each other, calculated as the proportion of possible triangles within the network that exist. |
| Dyad reciprocity (global) | The proportion of all edges in the network that were reciprocal. |
| Node characteristics | |
| Neighbors | The total number of unique participants an individual was linked to within the full network. |
| Indegree | The total number of contacts with a given individual *reported by other study participants* on the weekly surveys. |
| Outdegree | The number of contacts with other participants *an individual reported* across the study period. |
| Dyad reciprocity (node) | The proportion of an individual's reported contacts that were also reported by their neighbors. |
| Closeness | A measure of the average distance (i.e., number of edges) between a node and all other nodes in the network. |
| Clustering coefficient | A measure of how many of a node's neighbors are connected to each other; an individual measure of transitivity. |
| Network exposure - drinking | The proportion of a participant's neighbors that were categorized as drinkers (including binge drinkers). |
| Network exposure – binge drinking | The proportion of a participant's neighbors that were categorized as binge drinkers. |
| Dyad characteristics | |
| Dyad reciprocity (dyad) | The proportion of multi-edges that were reciprocal, within a given dyad (pair of connected nodes), |

**Table 2.2** Demographic and Social Network Characteristics of Year 1 (2012) Participants by Reported Baseline Alcohol Use (N=481).

| **Variable** | **N (%) or Mean (SD)**[a] | | |
|---|---|---|---|
| | **Drinker** | **Non-drinker** | ***P value***[b] |
| N | 163 (33.89) | 318 (66.11) | |
| Age | | | |
| <21 | 139 (85.28) | 296 (93.08) | <0.01* |
| 21+ | 24 (14.72) | 22 (6.92) | |
| Gender | | | |
| Female | 94 (57.67) | 182 (57.23) | 0.93 |
| Male | 69 (42.33) | 136 (42.77) | |
| Academic year | | | |
| Freshman | 106 (65.03) | 190 (59.75) | 0.07 |
| Sophomore | 26 (15.95) | 79 (24.84) | |
| Junior or higher | 31 (19.02) | 49 (15.41) | |
| Race | | | |
| White | 135 (82.82) | 209 (66.99) | <0.01* |
| Asian | 17 (10.43) | 55 (17.63) | |
| Black or African American | 6 (3.68) | 29 (9.29) | |
| Multi-racial or other race[c] | 5 (3.07) | 19 (6.09) | |
| Network characteristics | | | |
| Neighbors | 6.45 (6.80) | 5.34 (5.86) | 0.06 |
| Indegree | 18.34 (22.37) | 14.57 (18.17) | 0.05* |
| Outdegree | 19.10 (28.14) | 15.69 (24.34) | 0.17 |
| Dyad reciprocity (node) | 0.27 (0.23) | 0.29 (0.25) | 0.36 |
| Clustering coefficient | 0.46 (0.36) | 0.40 (0.35) | 0.09 |
| Network exposure–drinking[d,e] | 0.42 (0.29) | 0.31 (0.28) | <0.0001* |
| Network exposure–binge drinking[e,f] | 0.27 (0.27) | 0.16 (0.23) | <0.0001* |

Abbreviation: SD, standard deviation.
*P<0.05

[a] Numbers may not sum to totals due to missing data.

[b] P-values calculated by Chi-squared or Fisher's Exact test for categorical variables or resampling tests for network variables.

[c] Native Hawaiian or other Pacific Islander, American Indian, Alaskan Native, or Multi-racial.

[d] The proportion of neighbors that were drinkers.

[e] Not calculated for individuals without any neighbors (i.e., in- and out-degree of 0).

[f] Proportion of neighbors that were binge drinkers.

**Table 2.3** Demographic and Social Network Characteristics of Year 2 (2013) Participants by Reported Baseline Alcohol Use (N=423).

| Variable | N (%) or Mean (SD)[a] | | |
|---|---|---|---|
| | **Drinker** | **Non-drinker** | ***P value*[b,c]** |
| N | 155 (36.64) | 268 (63.36) | |
| Age | | | |
| <21 | 121 (78.06) | 245 (91.42) | <0.001* |
| 21+ | 34 (21.94) | 23 (8.58) | |
| Gender | | | |
| Female | 90 (59.21) | 168 (63.40) | 0.40 |
| Male | 62 (40.79) | 97 (36.60) | |
| Academic year | | | |
| Freshman | 85 (55.92) | 147 (55.26) | 0.02* |
| Sophomore | 29 (19.08) | 77 (29.95) | |
| Junior or higher | 38 (25.00) | 42 (15.79) | |
| Race | | | |
| White | 111 (74.50) | 162 (62.55) | 0.02* |
| Asian | 22 (14.77) | 51 (19.69) | |
| Black or African American | 7 (4.70) | 33 (12.74) | |
| Multi-racial or other race[c] | 9 (6.04) | 13 (5.02) | |
| Network characteristics | | | |
| Neighbors | 9.34 (8.64) | 6.87 (6.60) | <0.01* |
| Indegree | 33.34 (34.73) | 21.87 (23.30) | <0.0001* |
| Outdegree | 37.44 (52.09) | 24.32 (35.58) | <0.01* |
| Dyad reciprocity (node) | 0.26 (0.20) | 0.28 (0.24) | 0.45 |
| Clustering coefficient | 0.47 (0.32) | 0.40 (0.32) | 0.03* |
| Network exposure–drinking[d,e] | 0.44 (0.28) | 0.26 (0.26) | <0.0001* |
| Network exposure–binge drinking[e,f] | 0.20 (0.21) | 0.11 (0.16) | <0.0001* |

Abbreviation: SD, standard deviation.
*P<0.05
[a] Numbers may not sum to totals due to missing data.
[b] P-values calculated by Chi-squared or Fisher's Exact test for categorical variables or resampling tests for network variables.
[c] Native Hawaiian or other Pacific Islander, American Indian, Alaskan Native, or Multi-racial.
[d] The proportion of neighbors that were drinkers.
[e] Not calculated for individuals without any neighbors (i.e., in- and out-degree of 0).
[f] Proportion of neighbors that were binge drinkers.

**Table 2.4** Demographic and Social Network Characteristics of Year 1 (2012) Participants by Reported Alcohol Consumption Volume, Among Baseline Drinkers (N=163).

| **Variable** | **N (%) or Mean (SD)[a]** | | |
|---|---|---|---|
| | **Binge Drinkers** | **Moderate Drinkers** | ***P value*[b]** |
| N | 90 (56.25) | 70 (43.75) | |
| Age | | | |
| <21 | 82 (91.11) | 54 (77.14) | 0.01* |
| 21+ | 8 (8.89) | 16 (22.86) | |
| Gender | | | |
| Female | 49 (54.44) | 42 (60.00) | 0.48 |
| Male | 41 (45.56) | 28 (40.00) | |
| Academic year | | | |
| Freshman | 72 (80.00) | 32 (45.71) | <0.001* |
| Sophomore | 8 (8.89) | 17 (24.29) | |
| Junior or higher | 10 (11.11) | 21 (30.00) | |
| Race | | | |
| White | 77 (85.56) | 55 (78.57) | 0.59 |
| Asian | 7 (7.78) | 10 (14.29) | |
| Black or African American | 3 (3.33) | 3 (4.29) | |
| Multi-racial or other race[c] | 3 (3.33) | 2 (2.86) | |
| Network characteristics | | | |
| Neighbors | 6.04 (5.73) | 7.20 (7.98) | 0.29 |
| Indegree | 15.60 (17.16) | 22.44 (27.42) | 0.05* |
| Outdegree | 16.96 (17.98) | 22.69 (37.33) | 0.21 |
| Dyad reciprocity (node) | 0.25 (0.21) | 0.30 (0.25) | 0.19 |
| Clustering coefficient | 0.44 (0.29) | 0.49 (0.36) | 0.37 |
| Network exposure – drinking[d,e] | 0.44 (0.29) | 0.39 (0.28) | 0.27 |
| Network exposure – binge drinking[e,f] | 0.28 (0.27) | 0.26 (0.27) | 0.75 |

Abbreviation: SD, standard deviation.

*P<0.05

[a] Numbers may not sum to totals due to missing data. Baseline drinkers without detailed drinking data could not be categorized into consumption categories (N=3).

[b] P-values calculated by Chi-squared or Fisher's Exact test for categorical variables or bootstrapping for network variables.

[c] Native Hawaiian or other Pacific Islander, American Indian, Alaskan Native, or Multi-racial.

[d] The proportion of neighbors that were drinkers.

[e] Not calculated for individuals without any neighbors (i.e., in- and out-degree of 0).

[f] Proportion of neighbors that were binge drinkers.

**Table 2.5** Demographic and Social Network Characteristics of Year 2 (2013) Participants by Reported Alcohol Consumption Volume, Among Baseline Drinkers (N=155).

| **Variable** | **No. (%) or Mean (SD)**[a] | | |
|---|---|---|---|
| | **Binge Drinkers** | **Moderate Drinkers** | ***P value***[b] |
| No. | 75 (50.34) | 74 (49.66) | |
| Age | | | |
| <21 | 66 (88.00) | 49 (66.22) | <0.01* |
| 21+ | 9 (12.00) | 25 (33.78) | |
| Gender | | | |
| Female | 42 (56.00) | 47 (64.38) | 0.30 |
| Male | 33 (44.00) | 26 (35.62) | |
| Academic year | | | |
| Freshman | 48 (64.00) | 34 (46.58) | <0.01* |
| Sophomore | 17 (22.67) | 11 (15.07) | |
| Junior or higher | 10 (13.33) | 28 (38.36) | |
| Race | | | |
| White | 55 (73.33) | 54 (77.11) | 0.76 |
| Asian | 13 (17.33) | 8 (11.43) | |
| Black or African American | 3 (4.00) | 4 (5.71) | |
| Multi-racial or other race[c] | 4 (5.33) | 4 (5.71) | |
| Network characteristics | | | |
| Neighbors | 7.68 (7.57) | 11.53 (9.33) | <0.01* |
| Indegree | 25.39 (27.09) | 43.43 (39.53) | <0.01* |
| Outdegree | 31.07 (46.26) | 46.46 (57.44) | 0.31 |
| Dyad reciprocity (node) | 0.26 (0.18) | 0.29 (0.22) | 0.53 |
| Clustering coefficient | 0.44 (0.33) | 0.51 (0.30) | 0.22 |
| Network exposure–drinking[d,e] | 0.43 (0.31) | 0.45 (0.24) | 0.67 |
| Network exposure–binge drinking[e,f] | 0.23 (0.26) | 0.18 (0.14) | 0.20 |

Abbreviation: SD, standard deviation.

*$P<0.05$

[a] Numbers may not sum to totals due to missing data. Baseline drinkers without detailed drinking data could not be categorized into consumption categories (N=3).

[b] P-values calculated by Chi-squared or Fisher's Exact test for categorical variables or bootstrapping for network variables.

[c] Native Hawaiian or other Pacific Islander, American Indian, Alaskan Native, or Multi-racial.

[d] The proportion of neighbors that were drinkers.

[e] Not calculated for individuals without any neighbors (i.e., in- and out-degree of 0).

[f] Proportion of neighbors that were binge drinkers.

**Table 2.6** Social Network Properties and Structure of the Year 1 (2012) and Year 2 (2013) Social Networks, Containing All Enrolled Undergraduate Participants.

| Characteristic | Year 1 | Year 2 |
|---|---:|---:|
| Nodes | 584 | 590 |
| Multi-edges | 6306 | 9957 |
| Multi-edges per dyad (edge weight)[a] | 4.27 (2.74) | 5.34 (3.36) |
| Dyad reciprocity (global) | 1923 (30.49) | 2832 (28.44) |
| Density | 0.003 | 0.004 |
| Transitivity[b] | 0.54 | 0.58 |
| Isolates[c] | 94 (16.09) | 86 (14.57) |

[a] Mean (Standard deviation)

[b] Not defined for multigraphs; computed by flattening multi-edges into single edges.

[c] N (%)

**Table 2.7** Social Network Density and Transitivity of Year 1 (2012) and Year 2 (2013) Alcohol Use Sub-Networks. Each sub-network contains only participants in a given baseline alcohol consumption category.

| Sub-Network | Density | Transitivity |
|---|---:|---:|
| **Year 1** | | |
| Drinker | 0.0010 | 0.67 |
| Non-drinker | 0.0006 | 0.48 |
| Moderate drinker | 0.0011 | 0.76 |
| Binge drinker | 0.0011 | 0.72 |
| **Year 2** | | |
| Drinker | 0.0018 | 0.77 |
| Non-drinker | 0.0009 | 0.47 |
| Moderate drinker | 0.0031 | 0.83 |
| Binge drinker | 0.0012 | 0.54 |

## 2.7 Figures

**Figure 2.1** Number of reported drinks consumed in by undergraduate students in Year 1 (2012) and Year 2 (2013) by day of the week, among participants who reported alcohol use at baseline. (Year 1: N=163; Year 2: N=155). Less than 10% of drinkers reported any alcohol consumption on Sunday, Monday, Tuesday, and Wednesday.

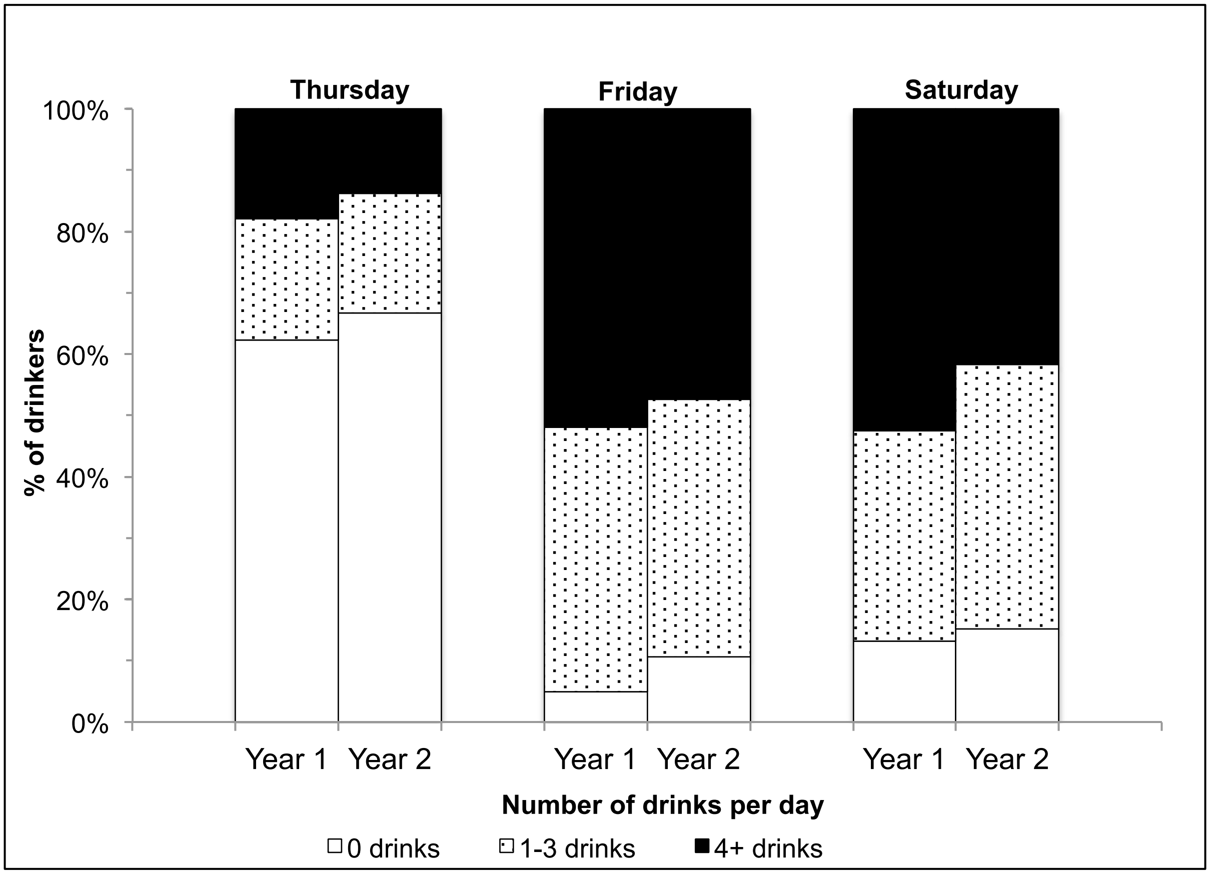


**Figure 2.2** Year 1 (A; 2012) and Year 2 (B; 2013) social networks of undergraduate students. Nodes (i.e., participants) are represented as circles, colored according to alcohol use; participants who declined to report their alcohol use are represented by the smaller gray circles. Each directed edge (i.e., line) represents a reported contact between 2 participants on a weekly survey, with an arrow pointed towards the reported contact. Reciprocal contacts reported by both participants are represented by double arrows. Edges are colored according to the alcohol use of the 2 connected nodes.

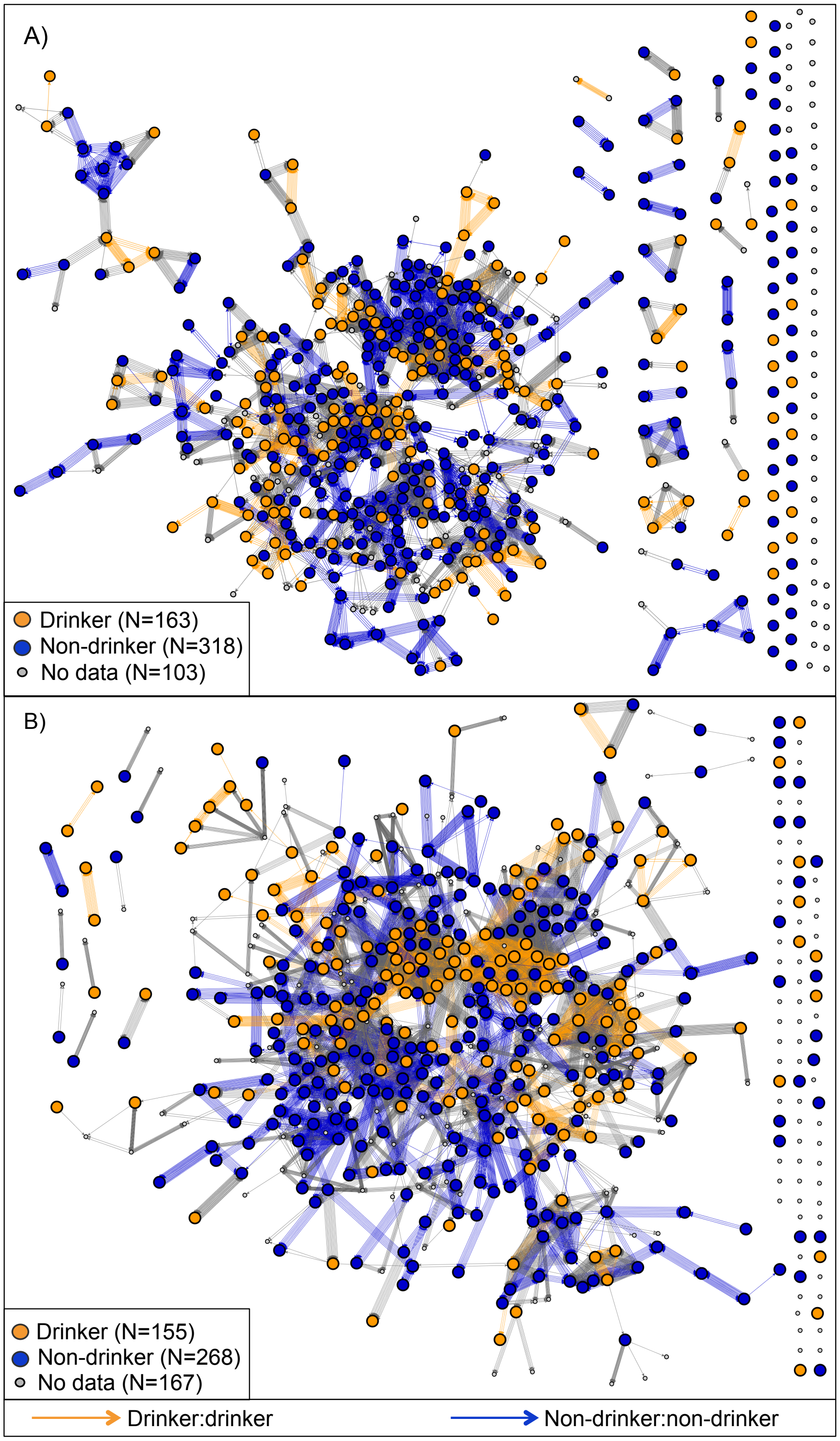


**Figure 2.3** Mean multi-edges and dyad reciprocity (mean (standard deviation)) between dyads of drinkers and non-drinkers in the Year 1 (A; 2012) and Year 2 (B; 2013) social networks of undergraduate students. Undirected multi-edges between pairs of nodes represent the edge's weight, or the frequency of contacts, i.e., the mean number of weeks a contact was reported (reciprocal and/or un-reciprocal) across the study periods (Year 1: 8 weeks; Year 2: 10 weeks). The mean proportion of reciprocal edges within dyads is also listed. Mean multi-edge weight and reciprocity between pairs of non-drinkers and between drinkers and non-drinkers were compared to the mean between drinkers by bootstrapping.

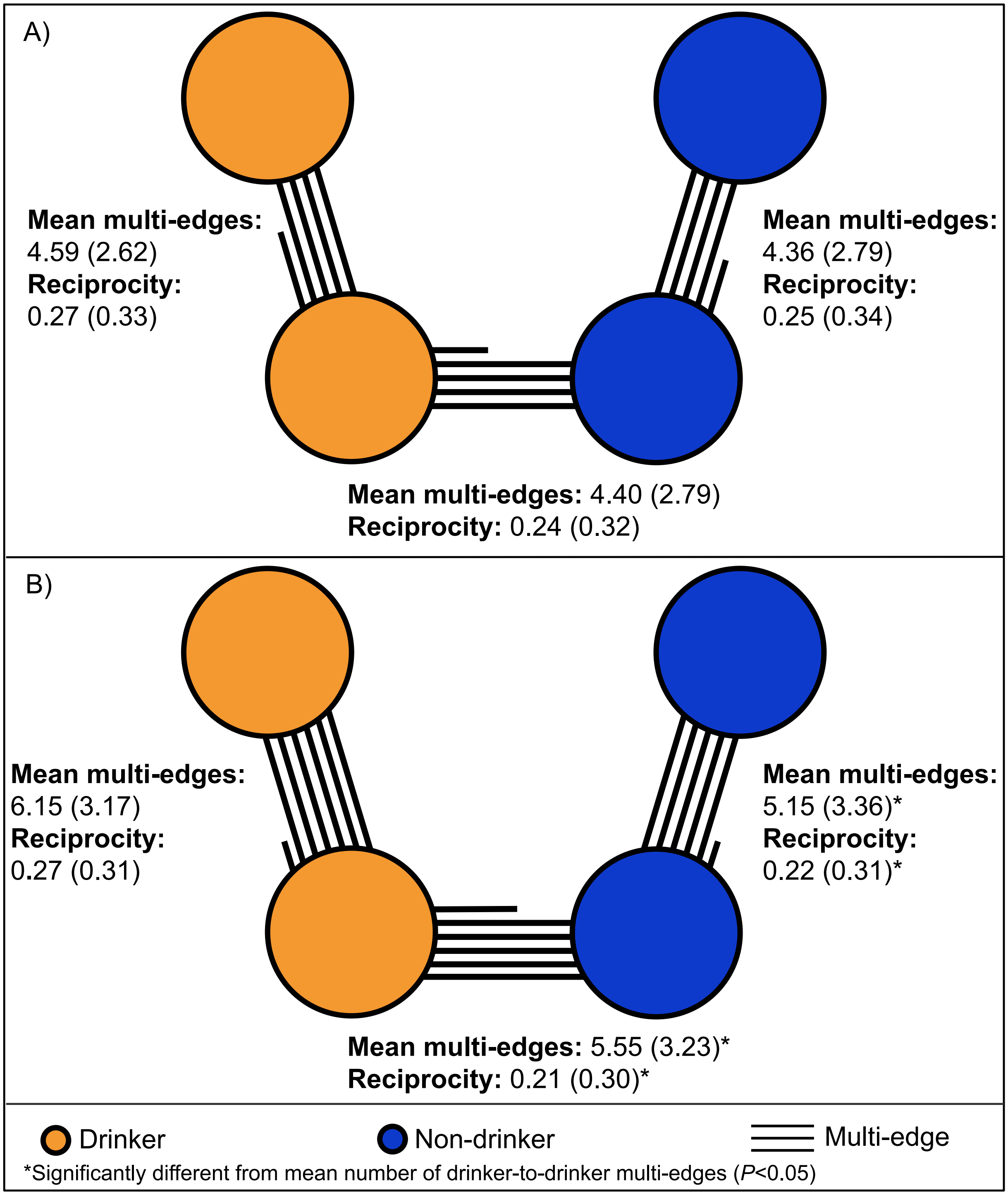
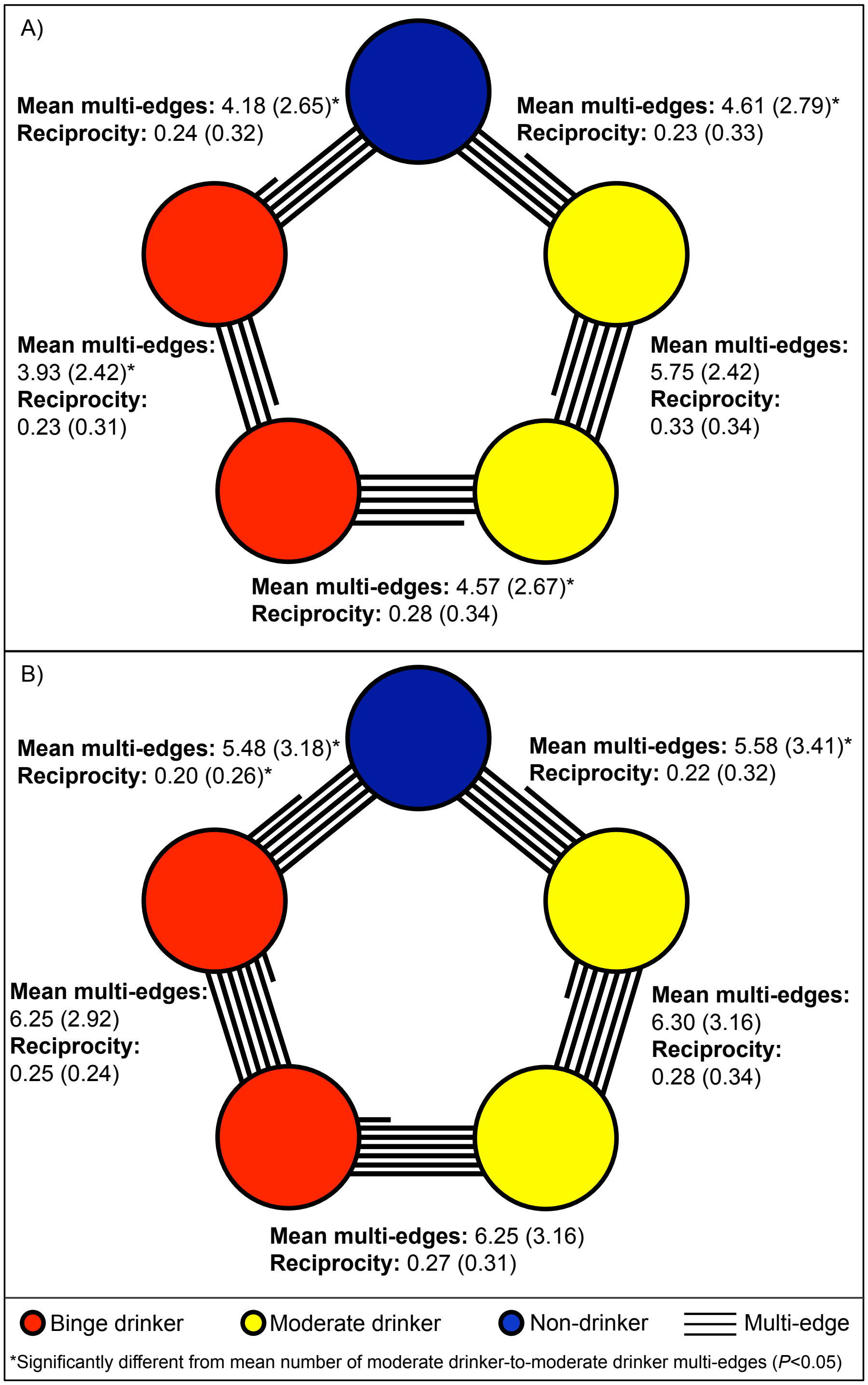


**Figure 2.4** Mean multi-edges and dyad reciprocity (mean (standard deviation)) between dyads of binge, moderate, and non-drinkers in the Year 1 (A; 2012) and Year 2 (B; 2013) social networks of undergraduate students. Undirected multi-edges between pairs of nodes represent the edge's weight, or the frequency of contacts, i.e., the mean number of weeks a contact was reported (reciprocal and/or un-reciprocal) across the study periods (Year 1: 8 weeks; Year 2: 10 weeks). The mean proportion of reciprocal edges within dyads is also listed. Bootstrapped resample was used to compare mean multi-edge weight and reciprocity between dyads of moderate drinkers and other dyad types.



*Significantly different from mean number of moderate drinker-to-moderate drinker multi-edges ($P<0.05$)

# Chapter 3

# Comparing Sensor and Reported Empirical Networks of College Students

## 3.1 Introduction

Social mixing and contact patterns have become increasingly important in infectious disease, social contagion, and health behavior transmission research (1,8,199,246) and the ability to accurately measure and quantify contact patterns is essential to building realistic transmission models and targeting interventions (13). However, contact patterns relevant to respiratory disease transmission are particularly challenging to quantify and measure (13,247), as transmission-relevant contacts may potentially be of short duration and may not always occur between close social ties that would be captured in self-reported survey data. Currently, real-world (as opposed to theoretical or homogenous mixing assumptions) social network data collection methods include self-report/diaries, contact sensors or “motes,” and direct observation, each with their own advantages and drawbacks, and each resulting in different observed networks with unknown validity. Individual contact patterns are influenced by a number of factors, including age, socio-economic status, and psychosocial factors, in addition to environment and temporality (e.g., urban versus rural, summer versus winter) (154,197,199,248–253). Additionally, unlike many static or single-time-point network measurement approaches, real-world networks are dynamic, with wide day-to-day and hypothetically, longer term, variations in interaction duration and distance that can indicate distinct types of social ties and yield different health outcomes (184,251,253–257). Thus, not only is it important to accurately measure individual contact patterns but also patterns and stability in dynamic, longitudinal networks.

There is a growing body of literature on theoretical social network data measurement error and its potential impacts on network characteristic estimates (164,209,258) and transmission models (182,195,259). Indeed, many inferences made about social and transmission dynamics rely on observed social networks and characteristics for which “we cannot estimate robustness or uncertainty” (205). However, contact surveys or diaries are currently the commonly used method for collecting empirical social contact data (171). Although self-reported contact data may more closely capture real-world social mixing patterns than homogenous mixing and theoretical networks, the extent to which these methods capture the underlying “true” social network and in particular, contacts most relevant to airborne (or other modes of) disease transmission is unclear (172). A number of social network studies have found that longer duration, higher frequency, and/or close relationship-contacts represent the majority of self-reported contacts (177,184,187,198,260,261). These self-reported contacts may capture a sense of the contact’s social importance to the participant (which can also have important health implications, e.g., behavior propagation and social influence), but may not give a useful representation of the set of contacts that facilitate transmission. For example, proximity and duration of contacts can be more important in direct and/or airborne transmission than other variables, such as the strength or type (e.g., good friend versus coworker) of the relationship (13,56). Individuals may have regular, transmission-relevant contacts that they do not personally know who would not be captured in survey, and individual’s recalled interaction durations are often incorrect or missed completely (184,195,262), with difficult to measure biases (11,161,260). Generally, self-reported contacts may be more appropriate for analyzing network selection and influence related to ideas and behaviors than for examining disease transmission, as self-reported contacts may have biases that could impact estimates of critical disease parameters (184).

Researchers are thus increasingly using wearable sensors to objectively capture dynamic social networks across a variety of settings (184–190,263). However, sensor contact-detection methods have a number of practical challenges (182,191,192), including generating large amounts of data that can be difficult to organize, parse, store, and analyze (174), as well as participant privacy concerns (193). Additionally, sensors only capture contacts between participants wearing active sensors, and many sensor types (e.g. Bluetooth) may record un-meaningful contacts, such as between participants separated by a wall (13). While survey methods capture fewer contacts, they have the advantage of being comparatively easier studies to run, can be run pro- or retrospectively, and can collect egocentric contact data as well as complete, within-study network data (13). However, there are known drawbacks to contact diaries including measurement error and recall issues, especially in longitudinal studies, where contact under-reporting may occur in order to avoid additional questions (“panel conditioning”) (264,265). Additionally, contact diary or survey format (e.g., phone survey, web or paper-based diary) can impact the number of contacts participants report (171,264,265). As such, comparing networks collected via contact surveys and sensors is a key step in determining the veracity of and stability of empirical networks observed via different data collection methods. A handful of previous short-term sensor and diary-based network comparisons have been conducted in middle and high schools (185,187), as well as at an academic conference (194), with little consistency in the observed agreement between measurement methods. To our knowledge, there have not been any comparison studies of “real-world” networks, i.e., those that occur outside of a closed environment, like a school or conference, nor have there been any studies examining longer-term trends in comparability between methods. Given the significant temporal, contextual, and population variability in contact patterns (154,197,199,248–253), we note there continues to be a significant gap in our understanding of the comparability of observed sensor and survey-based social networks, especially those across different populations and over time.

In this study we have attempted to address this gap by comparing longitudinal social networks of undergraduate students using web-based contact surveys and high resolution sensor-based contact data. Indeed, college student social networks present a unique opportunity for network measurement and comparison. Not only a somewhat captive, and therefore unlikely to be lost to follow-up, population (during the school year), we posit that college students’ lives can be seen as microcosms of multiple types of real-world networks; classrooms with a fairly constant set of contacts, close-proximity housing and living quarters, and travel on and off-campus may approximate schools, households, communities and workplaces, respectively. We directly compared and contrasted 8 weeks of self-reported weekly contact data with contact data recorded by a Bluetooth-based proximity contact collection app, “iEpi” (266). Data for this study came from the eX-FLU social network study, a social network study examining a randomized intervention for isolating respiratory infection cases in a university setting (174), and from the iEpi eX-FLU sub-study. In this analysis, we characterize differences and similarities in the networks collected via these two data collection methods, as well as specific features of objectively recorded contacts, i.e., recorded by iEpi, that make these contacts more likely to also be captured via self-report (e.g., duration). The overall aims of this study of college student social networks were to: 1) compare social network measurement methods with varying levels of granularity in order to determine the comparability of data collected with two social network measurement tools; 2) identify the strengths and limitations of each approach in terms of the characteristics of captured contacts and the completeness of the measured networks; and 3) characterize observed changes in longitudinal contact patterns in order to estimate the extent and impact of declining data quality and participant engagement for sensor and survey-based network measurement.

## 3.2 Methods

### 3.2.1 iEpi recruitment

Data for these analyses come from the iEpi sub-study of the eX-FLU study, a social network study conducted in a large public university in the Midwestern United States during the winter 2013 semester (January 18-May 4, 2013). Detailed methods and protocols for the main eX-FLU and iEpi sub-study can be found in Aiello et al. 2016 (174). At the beginning of the winter semester, a sub-set of eX-FLU participants were invited to join the iEpi sub-study. Potential iEpi participants were identified and enrolled with the following protocol: first, the network was divided into modularity-based communities, using Girvan and Newman's recursive edge-deletion algorithm (267). Participants were then randomly selected from the largest community to receive an iEpi invitation. Waves of invitations were sent to random participants in successively smaller communities until the pre-determined population size was reached (N=103). Study staff passed out smartphones pre-loaded with the iEpi social network program at information sessions on January 28 and February 4 (the second and third weeks of parent study period), with an additional 'last chance' session on February 12th (parent study week 4); the majority of phones were deployed and recording data before the fourth week of the parent study (82.52%), and all phones were deployed and activated by week 6 (see Table 3.1 for iEpi activation numbers and dates). Participants were asked to carry the smartphones loaded with iEpi whenever on campus, from the time they received the phones from study staff through the end of the 10-week study period, excluding spring break (March 1-7). Participants could use the provided smartphone as their personal phone or as an additional device depending on personal preference, and were able to keep the provided smartphones at the study's conclusion, as a participation incentive.

### 3.2.2 Data collection

*iEpi data collection*

The iEpi app's specifications have been previously described in detail (266). Briefly, the app uses Bluetooth detection to identify proximity contacts (up to 5-10 meters) at 5-minute intervals beginning when the phone was powered on and the app activated. The app records the contact's unique identifier, and the contact date and time. Proximity contacts could be recorded by one or both iEpi apps. Notably, Bluetooth signals can transmit through walls and other physical barriers, so an iEpi recorded proximity contact did not necessarily indicate a true 'face-to-face' contact event. Conversely, iEpi does not record data when the phone is powered off, so the absence of a recorded proximity contact does not necessarily indicate a lack of face-to-face contact between two iEpi participants. Additionally, participants could "snooze" iEpi at any time for up to 12 hours, in 30-minute increments, during which iEpi did not record contact data and could not be 'seen' by other iEpi devices.

*Reported contact data collection*

Weekly web-based contact surveys were sent to all study participants during the 10 week study period, on which they were asked to report all other (parent study) participants with whom they had face-to-face contact during the previous week. Links to each weekly survey were sent on Fridays and responses were locked on the following Mondays (see Appendix A.1 for the weekly contact survey). Each participant selected an unlimited number of contacts, from a list of likely contacts (pre-populated with previously reported contacts) and/or by searching for study participant contacts by name/email. Data was de-identified using randomly assigned study IDs prior to analysis.

### 3.2.3 Data processing, matching, and network construction

In order to compare the weekly survey and iEpi sensor data collection methods, the time frames and sub-populations of the surveys and iEpi were matched in order to create networks for each study week. In other words, a network was constructed for each week of the study that contained identical nodes and all edges between them that were either captured by iEpi or reported on a given week's contact survey. Aggregated networks for each collection method, containing edges from all weekly networks were also constructed. A flowchart of the data processing, contact matching, and network construction process can be found in Figure 3.1. Raw iEpi contact data was aggregated and processed using the following set of steps/rules: First, contacts recorded by either member of a dyad were ordered according to date and time, with the first recorded contact on a given date considered the contact event's 'start time.' Although the app's recording frequency was 5 minutes, we considered subsequent contacts recorded by either participant's app within 10 minutes of the previous recorded contact to be a continuation of contact. This allowance was made to account for missed contacts by the Bluetooth sensor and/or brief separations within dyads that occurred at iEpi recording times but did not functionally end the contact, e.g., an individual going to a restroom and returning. When a gap between recorded contacts was greater than 10 minutes, the contact was considered to have ended 2.5 minutes after the last recorded time before the gap, and a new contact started at the first recorded event after the gap. Individual contact duration was then calculated using each contact event's start and end times. Contact events without a subsequent recorded contact within 10 minutes were assigned a duration of 2.5 minutes. Next, as the reported contact data granularity was at the week level, iEpi contacts within each study week were aggregated to match this time scale. Total proximity contact duration, or edge weight, between two participants, or dyads, was calculated as the sum of all of their unique contact durations within the study week, in hours.

To maintain congruence between nodes in each week's Reported ($RN_3$-$RN_{10}$) and iEpi ($iEN_3$-$iEN_{10}$) Network, nodes and their reported or iEpi recorded contacts were included in the weekly networks beginning with the first full study week after their iEpi app was activated, defined as a device's first recorded contact event. Note that weekly reported contacts were directed, whereas weekly iEpi contacts were undirected and weighted by the week's proximity contact duration (i.e., the sum of all individual recorded proximity contact durations in a given study week). However, within the weekly reported contacts, there was the possibility that a contact could have been reported by both members of a dyad in a given week. In this case, a single, reciprocal edge was drawn between the pair. Contacts that were reported by only one member of a dyad were identified as unreciprocal edges.

Aggregate networks ($RN_a$, $iEN_a$) for each measurement method were also constructed by combining all 8 weekly networks within each method. Edges in these networks are undirected and weighted by the number of weekly networks within each method that an edge was present. The directionality of weekly edges was not considered when calculating edge weights for the aggregated $RN_a$.

Network construction, as well as network data processing and matching, was done using the Python (2.7) NetworkX package (218).

### 3.2.4 Analysis

In order to compare reported versus iEpi/sensor contacts overall, as well as by study week, individual networks for each (parent) study week that contained at least one iEpi contact (weeks 3-10) were constructed. The resulting networks generated by each measurement method were compared pairwise by week across measurement methods. We hypothesized that iEpi proximity contacts with longer cumulative weekly duration would be reported more frequently on weekly surveys than shorter cumulative weekly duration contacts, so we constructed additional iEpi networks by restricting $iEN_3$-$iEN_{10}$ to edges with weekly durations (d) greater than 2.5 minutes, 0.25, 0.50, or 1 cumulative hours of recorded contact per week. Aside from 2.5 minutes of proximity contact per week, which was the minimum possible recorded contact duration that we set for iEpi contacts (corresponding a single recorded proximity contact between two participants), these thresholds were selected by pooling all weekly contact durations and dividing them into relatively equal groups with logical end points (see Figure 3.2). Networks were analyzed in Python (2.7) with the NetworkX package (218) and visualized in Visone 2.17 (Konstanz, Germany).

*Network descriptions and characteristics*

Each weekly and aggregate network was visualized with nodes at identical locations. Summaries of each network's construction and assessed network statistics and characteristics are shown in Table 3.2. For weekly and aggregate networks by measurement method, a number of network and node characteristics were assessed. Network density, transitivity, mean edge weight (i.e., total week contact duration; $iEN_3$-$iEN_{10}$ only), and dyad reciprocity (i.e., the proportion of all network dyads with reciprocal edges; $RN_3$-$RN_{10}$ only) were calculated and the numbers of nodes, edges, and isolates were counted. For aggregate networks ($RN_a$, $iEN_a$), mean edge weight, or number of weekly networks an edge appeared in, was also calculated. For nodes within a given network, degree was defined as a node's number of neighbors, or the number of other nodes connected to a particular node by an edge. In $RN_3$-$RN_{10}$, in- and out-degree represent the number of contacts a participant reported and the number of other participants who reported a given participant as a contact on a given weekly survey, respectively. For each network, mean degree, indegree ($RN_3$-$RN_{10}$ only), outdegree ($RN_3$-$RN_{10}$ only) were assessed.

*Degree Distributions and Sequences*

We plotted comparisons of $iEN_a$ and $RN_a$ degrees using a network status layout (in rows corresponding to $RN_a$ degree) to show potential individual differences in centrality between the two collection methods for the entire analyzed study period. Additionally, degree distributions were assessed for all networks and compared between data collection methods across study weeks 3-10 with bar and whisker plots, comparing each week's RN and iEN (all edges and all duration restricted networks; cumulative weekly duration ($d$) edge restrictions: $d > 2.5\ minutes$; $d \geq 0.25\ hours$; $d \geq 0.50\ hours$; $d \geq 1\ hour$). For each of the four duration restricted networks, a node's iEN degree was re-calculated and was defined as the number of neighboring nodes an individual was connected to with an iEpi edge with cumulative duration above the noted cutoff.

A summary of pairwise degree sequence dependence analyses across measurement methods is shown in Table 3.3. For each analyzed study week, we assessed the ordinal association between the $iEN_3$-$EN_{10}$ (for the full network and four duration restricted networks) and $RN_3$-$RN_{10}$ degree sequences (i.e., numbers of neighbors for each participant) with Kendall rank correlation coefficients (tau, τ; calculated using the stats package in the SciPy library (268)) and tested the hypothesis of statistical dependence via resampling. Kendall's tau coefficient assesses the degree of monotonic correlation, or degree of similarity, between observed, nonparametric sequences on the same set of individuals (269). We also assessed the association between indegree and outdegree sequences in $RN_3$-$RN_{10}$ and $iEN_3$-$EN_{10}$ sequences (all edges and four duration restricted networks), in order to determine if either indegree or outdegree distribution, or both, had higher levels of dependence with a given week's iEN degree sequence than the RN degree distribution. Hypothesis testing for Kendall's tau rank coefficients was run under the null hypothesis of independent distributions via resampling, using $\tau$ as the test statistic. Resampling was conducted by shuffling the set of $RN_3$-$RN_{10}$ degree values for all nodes to create new rank orders, or paired values of $iEN_3$-$iEN_{10}$ and $RN_3$-$RN_{10}$ degrees for each node. Upon resampling, the test statistic,$\tau$, was recalculated. Resamples were performed 10,000 times for each comparison, and the P-value represents the proportion of resampled coefficients as or more extreme than the original observed value.

*Network overlap and agreement*

Edges were matched between $iEN_3$-$iEN_{10}$ and $RN_3$-$RN_{10}$ by week, to crudely assess the amount of network overlap or congruence, i.e., the extent to which proximity contacts recorded by iEpi coincided with reported face-to-face contacts. Edges were matched overall by identifying dyads (i.e., pairs of participants connected by a reciprocal, unreciprocal, or undirected edge) that were present in both RN and iEN in any given study week. We then assessed the recorded cumulative weekly contact duration, or edge weight, of iEpi edges that did and did not have corresponding present edges in the same week's RN and determined the distribution and mean edge weight of iEN edges with and without corresponding RN edges. These edge weight distributions were visualized bar and whisker plots in order to characterize differences in cumulative weekly duration between iEpi recorded edges that were and were not reported by participants across the study period.

As there is not a standard method for comparing and quantifying agreement of two or more networks with identical or overlapping boundaries, (potentially) different sets of edges, different reporting/sensing methods, and no assumptions about any network's alignment with a 'true' network, we adapted the traditional Cohen's Kappa ($\kappa$) test of inter-rater agreement to assess global agreement between pairs or groups of networks. While we used undirected versions of the networks in these analyses, the method could be used for directed networks with a directed list of possible dyads as the comparison items; the caveat being that the networks being compared must both be treated as directed or undirected.

The basic framework for our formulation of a network $\kappa$ for two networks, A and B, is as follows. We generated a list of nodes as the union of the nodes in Network A and Network B. Using the potential dyads from this list as the agreement items, the methods were said to agree on a given item if a potential dyad was either present or absent in both the iEpi and Reported networks. If a dyad was present in one but not the other, this was considered disagreement on an item. The standard formula for Kappa is (270):

$$\kappa = \frac{p_0 - p_e}{1 - p_e}$$

where we adapt the standard $\kappa$ as follows: $p_0$ is the observed agreement between the two (or more) networks and $p_e$ is the expected agreement that would occur by chance, based on the number of edges in each network. Agreement is defined as the fraction of possible edges upon which both networks agree (i.e., both networks include a given edge/dyad *or* both networks do not include a given edge/dyad).

More formally, let $N_1, \dots, N_p$ be a collection of networks (e.g., resulting from different measurement methods, or measured at different times, or simply distinct networks of interest). In this analysis we consider $p = 2$, where we examine networks constructed from data collected with two distinct measurement methods (iEpi and Reported). For each network $N_k$, let $N_k = (V_k, E_k)$, where $V_k$ is a set of vertices and $E_k$ is a set of edges, defined as vertex pairs, i.e., for $e_{ij} \in E_k$, we define $e_{ij} = (v_i, v_j)$, where $v_i, v_j \in V_k$. We indicate potential edges or dyads which do not appear in $E_k$ with a bar, i.e., taking $\bar{E}_k := \{e_{ij} | e_{ij} \notin E_k\}$. If all networks are undirected, we treat $e_{ij} = e_{ji}$, i.e., we treat each edge as a set rather than an (ordered) tuple; otherwise, we treat directed networks as including both $e_{ij}$ and $e_{ji}$ separately. For convenience, and to ensure $\kappa$ is well defined, we take the vertex set for all networks to be the union of all possible vertices, $V \coloneqq \bigcup_{k=1}^{p} V_k$, and redefine our networks as $N_k = (V, E_k)$. With this notation established, $p_0$ is given by:

$$p_0 = \frac{\bigcap_{k=1}^{p} E_k + \bigcap_{k=1}^{p} \bar{E}_k}{E^*}$$

.

Where $E^*$ is the total possible edges, given by $(|V| \cdot |V - 1|)/2$ for a collection of undirected networks and $|V| \cdot |V - 1|$ for a collection including directed networks. Next, we note that the probabilities of a given edge appearing or nor appearing in a given network $N_k$ are given by

$|E_k| \,/\, E^*$ and $|\bar{E}_k| \,/\, E^*$ respectively (noting that the edge probability is also commonly called the network density). Then the overall agreement by chance is:

$$p_e = \prod_{k=1}^{p} \frac{|E_k|}{E^*} + \prod_{k=1}^{p} \frac{|\bar{E}_k|}{E^*}$$
$$= \prod_{k=1}^{p} \frac{|E_k|}{E^*} + \prod_{k=1}^{p} \left(1 - \frac{|E_k|}{E^*}\right).$$

A table summarizing all pairwise Kappa-based network agreement analyses is shown in Table 3.4. We assessed agreement between weekly network pairs and between the two aggregate networks. We also compared the $RN_3$ - $RN_{10}$ to its corresponding $iEN_{3,\ d<[cutoff]}$ - $iEN_{10,\ d<[cutoff]}$ (weekly duration restricted iEN), to determine if overall agreement between reported contacts and proximity contacts increased as we raised the included recorded cumulative weekly duration by the iEpi app.

## 3.3 Results

### 3.3.1 Study population

Study population demographics and distributions were previously assessed in Aiello et al. 2016 (174). The majority of the 103 iEpi participants were White (61.4%), female (66%), and freshman (56.3%); the mean age of iEpi participants was 18.9 (SD: 1.1) years (see Table 3.5).

### 3.3.2 Network data, descriptions, and characteristics

Network visualizations of the weekly and aggregated networks are shown in Figure 3.3. Descriptive data and characteristics of $iEN_3$-$iEN_{10}$ and $RN_3$-$RN_{10}$ are shown in Tables 3.6 and 3.7, and for the Aggregate networks ($iEN_a$ and $RN_a$), Table 3.8. iEpi sub-study participants reported 1076 face-to-face contacts with each other across the full parent study period, representing a total of 614 dyads. Of these dyads, 545 occurred were included in these analyses (contact events sensed prior to the first full study week after iEpi deployment for a given participant were excluded), with an overall dyad reciprocity of 0.43 across study weeks 3 through 10. When aggregated, the Reported network contained 121 dyads; each dyad was reported by one or both participants on an average of 4.50 (SD: 2.46) weekly surveys. The iEpi app recorded 453,281 proximity contact events between participants across the full deployment period, which corresponded to 358,369 contacts and 6,671 weekly proximity contacts. Of these, 3,565 weekly, duration-weighted iEpi contacts were analyzed across parent study weeks 3 through 10. In aggregate, the iEpi app recorded 1,977 dyads across the analyzed deployment period, with a mean weight of 1.80 (SD: 1.35).

The number of nodes in each network corresponds to the number of participants who had received and activated their iEpi device in a previous study week, e.g., 59 participants activated received and activated their iEpi device before study week 3, and were therefore included in the week 3 networks. Although the $RN_a$ had 24 isolates (23.3%), individual networks had higher numbers of isolates, or participants who did not have any reported contact in a given study week; approximately 35% of nodes in any given week were isolates in $RN_3$-$RN_{10}$ (range: 31.31-42.37%; see Table 3.6). The $iEN_a$ did not have any isolated nodes, as by definition all 103 nodes had at least 1 contact event recorded across the analyzed study period (as the first contact event is how we determined activation week). However, $iEN_3$-$iEN_{10}$ had increasing numbers of isolates over time, rising from 0 in week 3 to a high of 39 (37.86%) in week 10 (see Table 3.7).

Figures highlighting trends in global network and average node characteristics across the study period, and in aggregate, by data collection method are shown in Figure 3.4. For the most part, the week 3 networks were anomalous (particularly the $iEN_3$ network) to the rest of the study period in terms of density, transitivity, and mean degree. After week 3, density and transitivity remained fairly steady within $iEN_4$-$iEN_{10}$ and $RN_4$-$RN_{10.}$ While the $iEN_3$ had higher values of density and transitivity than the analogous $RN_3$, $iEN_4$-$iEN_{10}$ transitivity dropped precipitously below those observed in $RN_4$-$RN_{10}$, and remained approximately 50% lower than the $RN_4$-$RN_{10}$ transitivities for the remainder of the study period. $iEN_3$-$iEN_{10}$ densities also dropped rapidly between weeks 3 and 4 and continued to decline, albeit at a lower rate, for the rest of the study period, and the density of $iEN_{10\ was}$ slightly twice that of the density of $RN_{10}$. Based on the mean edge weights observed in the aggregated networks ($RN_a$, $iEN_{a;}$ see Table 3.8), we can also conclude that reported contact patterns were more consistent or stable across the analyzed study period than iEpi captured contacts. Each edge in the $RN_a$ was reported by one or both dyad members on an average of 4.50 (SD: 2.46) weekly surveys, whereas unique proximity contacts were only captured by iEpi in an average of 1.80 (SD: 1.35) weeks.

The two aggregated networks had very different global and average node characteristics (see Table 3.8). The $iEN_a$ had a higher density, due to many more edges (1,977 edges) than the $RN_a$ (121 edges), as well as a higher mean degree than the $RN_a$. However, the $RN_a$ was slightly more clustered than the $iEN_a$, with a higher observed transitivity.

### 3.3.3 Degree Distributions and Sequences

Figure 3.5 shows the correlation, or lack thereof between nodes' aggregate iEpi and Reported degrees. Although it appears that some nodes had similar aggregate iEpi and Reported degrees, based on visual inspection there does not appear to be a consistent or identifiable relationship between a node's degree in one network versus the other in the aggregated networks.

We also looked at degree distributions for both measurement methods for the eight weeks in the analyzed study period. Box and whisker plots showing iEpi degree distributions compared to Reported Network degree, indegree, and outdegree distributions by study week are shown in Figure 3.6. Across all considered study weeks, the distributions of iEpi degree had higher means and medians, as well as wider interquartile ranges, although the difference between iEpi and Reported distributions lessened over the analyzed weeks. Visually, it does not appear that there is greater concordance between iEpi degree distributions and the Reported outdegrees and indegrees than Reported overall degrees. This is further confirmed by the sequence dependence analyses results, which are shown in Figure 3.7. Overall, values of Kendall's rank correlation coefficient ($\tau$), or level of monotonic association, across different network measurement methods, ranges from approximately 0.04 to 0.20, with dependence levels across the three Reported Network degree types increasing slightly between weeks 3 and 10. Although a handful of values are statistically significant, indicating a level of dependence significantly different from 0, or complete independence, the values of $\tau$ for all 24 comparisons are too low to point to a meaningful relationship between degree (number of contacts) captured by the iEpi app and those reported by participants on weekly surveys.

However, concordance between the Reported degree distributions and those of the four cumulative weekly duration restricted iEpi networks (see Figure 3.8) was higher than the degree distributions from the full iEpi weekly networks, as the full iEpi weekly networks contain more edges than the reported networks. For each successive duration restriction, the iEpi degree distribution more closely resembles that of the Reported network in a given week, up to a point. For most weeks, it appears that restricted iEpi network edges to those that had cumulative proximity contact durations of at least one hour per week most closely resembled those of the Reported Networks. In other words, among the iEpi edge durations tested, participants' numbers of *reported* contacts (overall and for directional reporting) most closely resembled the numbers of people with whom participants' iEpi devices recorded at least one hour of proximity contact in a given week. Towards the end of the study period (weeks 9 and 10), there is a slight anomaly in the relationships, as the most restrictive iEpi network degrees were slightly below the Reported degrees, although this difference does not look meaningful.

Similarly, the level of dependence of the degree sequences, as measured by $\tau$, shows increasing concordance, or dependence, with each increasingly restrictive edge duration definition (see Figure 3.9). As expected based on the bar and whisker plots shown in Figure 3.8, there is increasing and significant agreement with increasingly higher duration proximity contact networks, with slight variation over time. The highest level of significant dependence between the Reported Network degree sequences and iEpi duration restricted Networks degree sequences was in week 6, with iEpi edges restricted to a cumulative weekly proximity contact duration of at least one hour. However, the even when restricted the proximity contacts to those with longer durations, the maximum value of $\tau$ we calculated was 0.35, which corresponds to "weak" agreement according to the generally accepted $\tau$ interpretation (thresholds) (269) and therefore this maximum calculated value many not represent a practically meaningful relationship between degree sequences across contact measurement methods. However, given the relative novelty of applying this statistic to degree sequences, the expected and threshold values of $\tau$ for an acceptable or high level of agreement is unknown. It should be noted that generally, regardless of contact duration, as measured by iEpi, there was a slight increase in dependence level across the study weeks, with week 10 having a Reported degree sequence that more closely resembled that of the week 10 iEpi Network with all edges.

### 3.3.4 Network overlap and agreement

Overall network overlap, or the numbers of edges or dyads that were present in both weekly iEpi and Reported Networks, overall and according to iEpi edge weight, i.e., cumulative weekly proximity contact duration, are shown in Table 3.9. Overall levels of overlap, or edge matching, by study week is shown in Figure 3.10. Generally, low percentages of proximity contacts captured by iEpi were also reported on weekly surveys. There were also noticeable numbers of contacts that participants reported on surveys that were not captured by iEpi. Importantly, the iEpi Networks increased the percentage of total network edges that were also present in a given week's Reported Network over the analyzed study weeks, rising from matching approximately 4% of network edges with corresponding reported edges in week 3 to approximately 12% in week 10. The opposite trend was seen in the Reported Networks; over the study period, the percentage of reported edges that were also captured by the iEpi app decreased from a high of 92% in week 3 to (although only 24 edges were present in this network) 39% by week 10. In other words, fewer reported contacts were captured by iEpi over the study period, but the amount of iEpi captured contacts that were also reported increased over the study period. This may be due in part to the decreasing numbers of edges in the iEpi networks as the study progressed.

There appears to be a relationship between cumulative weekly contact duration, as measured by iEpi, and the probability of that contact also being reported on a weekly survey. A pooled distribution of all iEpi edges' durations, as well as distributions by study week, were categorized by whether or not there was a corresponding edge in a given Reported Network shows the higher mean, median, and distribution of cumulative weekly contact duration between edges with and without matches in Reported Networks (see Figure 3.11). The mean cumulative weekly duration of all iEpi edges without corresponding reported edges was 3.80 hours/week (SD: 4.40) whereas the mean duration for edges that were also reported on a weekly survey was 0.62 hours/week (SD: 1.91). Across the study period, the distribution of cumulative weekly contact duration of iEpi edges without corresponding Reported edges were fairly normal (see Figures 3.11-3.12) whereas those with corresponding edges in a given week's Reported network had a the duration distribution for edge with corresponding reports on a weekly survey were skewed towards longer cumulative weekly durations. In other words, longer cumulative weekly duration contacts were more likely to be reported on a weekly survey than shorter cumulative weekly duration contacts. This appears to be especially true of iEpi contacts with a duration of 2.5 minutes, corresponding to a single recorded contact event; these contacts comprise very low percentages of iEpi contacts that were also reported, as shown in Figure 3.12.

These results are reinforced by our agreement analyses with the adapted Kappa statistic, as shown in Figure 3.13. When considering all edges in the weekly iEpi and Reported Networks, there is an upward trend in agreement over the study period, with week 10 reaching a kappa of approximately 0.20. Interestingly, the aggregate networks had much higher levels of agreement than all individual weeks, with a kappa value of 0.46, suggesting that over the course of the study, iEpi and Reported contacts that did not match in the opposite network in an early week were somewhat likely to be captured by the opposite data collection method in a subsequent week. However, even with the higher aggregate agreement, all network pairwise comparisons had a 'weak' level of agreement (270). Restricting the weekly iEpi Networks by cumulative weekly duration increases the overall level of network agreement across all weeks, with the largest difference found in week 3, where the complete iEpi Network had an agreement level with the Reported Network of less than 0.05 and the iEpi Network containing only edges with weights greater than or equal to 1 hour per week had an agreement level of 0.33, or approximately 10 times that of the unrestricted iEpi Network. Generally, each successive edge restriction increased the agreement slightly, with a small improvement after dropping edges that represented a single proximity contact event between two participants in a single week, and a larger increase in agreement when edges were restricted to 0.25 hours per week or more. These results mirror the cruder matching results described above, whereby restricting the iEpi Networks by cumulative weekly duration is, in practice, increasing the percentage of iEpi edges with a corresponding reported contact in a given week.

## 3.4 Discussion

The method of social network data collection affects the resulting measured network and the resulting conclusions or inferences. We described and compared two longitudinal series of social contact networks within a group of 103 undergraduate students with varied definitions of contacts and levels of granularity. We analyzed objectively collected proximity contact data recorded by a Bluetooth based app, iEpi, and subjective survey-based reported face-to-face contacts. Our analyses directly compared multiple facets of the resultant network data, including network structure, node centralities, degree distributions/sequences, and global network agreement. To this end, we developed a modified version of Cohen's Kappa coefficient to directly compare 2 or more networks containing overlapping sets of nodes. We investigated network similarities between sensor-recorded networks and participant-reported networks overall (i.e., across the full analyzed study period), as well as longitudinal trends in agreement, and the role of contact duration in network similarity using a mix of social network analyses and traditional epidemiological and statistical methods.

Comparisons between the network properties, degree distributions/sequences, and assessments of edge agreement and global network agreement between the Reported and iEpi networks all point to generally low overlap and agreement between the iEpi captured proximity contacts and participant reported face-to-face contacts. However, we observed that when iEpi networks were restrict to higher cumulative duration proximity contacts, overlap and agreement increased. Overall, even after aggregating contact event data into weekly contacts, the iEpi app recorded more contacts than the weekly contact surveys; the contacts recorded by iEpi but missed by the contact surveys were comprised mainly of shorter cumulative duration contacts. Surprisingly, in addition to this expected result, we also found that a large proportion of the weekly participant reported face-to-face contacts were not captured by the iEpi app. The degree distribution/sequences comparisons between collection methods and the Kappa results highlight the differences and lack of discernable relationship between iEpi and reported networks. Collectively, the results of these comparisons suggest that: 1) longer duration proximity contacts tend to have a higher likelihood of also being reported on a survey; 2) survey and sensor compliance and participation levels may decrease over time, calling into question the veracity and completeness of longitudinal social network data collected at later time points, and 3) neither sociocentric data collection method (web-based surveys and cellphone-based apps) yields complete contact data.

All contacts between the 103 study participants during the analyzed study period fall in one of four categories: 1) contacts captured by both methods (i.e., agreement); 2) contacts captured by iEpi but not reported on a weekly contact survey; 3) contacts reported on a weekly contact survey but not captured by iEpi; and 4) contacts occurred but were neither captured by iEpi or weekly contact surveys. The goal of these analyses was to describe and compare the data and networks in order to describe the characteristics of contacts that end up in each of the first three categories; the fourth category is inherently unmeasurable, although we would hope it represents as little of the "true" network as possible. As noted by Smieszek et al. (194), differences between survey and sensor contact data can be attributed to "true" differences in each method's definition of "contact," or measurement error by either method. In other words, we can attribute disagreement between compared networks to either differences in the definitions of contact employed by the two data collection methods, missing iEpi contact records corresponding to participant-initiated pausing of data collection or loss of cellphone battery power, and/or survey measurement error.

We can describe each method's contact definition in terms of type of contact and the duration granularity of the method's contact definition. The iEpi app's contact definition is proximity-based, i.e., Bluetooth detection of another iEpi device within approximately 3 meters. This proximity-based detection does not necessarily indicate that two participants (or devices) had to be in the same room or on the same floor of a building, however, as Bluetooth detection is possible across walls and through floors and ceilings. Additionally, the app attempts to detect other iEpi devices at 5-minute intervals. Therefore, if 2 iEpi participants were within 3 meters of each other for 5 minutes or more, it was likely that one or both of the devices would record the contact. However, if their close proximity to each other lasted less than 5 minutes, an iEpi record of the contact was dependent on the timing of both devices' recording schedule (i.e., the sampling frequency). Participants could also have neglected to bring their iEpi device with them on-campus, intentionally stop their iEpi app's data collection by "snoozing" the app, turning off the device's Bluetooth detection, or powering down the device; unintentional halting of data collection could also occur if the device ran out of power. In this case, any proximity contacts between an individual and other iEpi participants would not be recorded by the individual's app or any of their proximity contacts. We note then that iEpi data quality or completeness and the resulting measured network depended on all participants who may have come within 3 meters of each other having their devices on their person, turned on, and iEpi apps active at all times. In other words, while only one device needed to record a proximity contact for it to be represented in the measured network, both devices present needed to be on and active for the contact to be recorded.

The weekly contact surveys defined contact as "face-to-face contact" occurring during the previous week. In effect, the contact survey reports relied on participants' recognition of any and all other study participants by name, and their ability to recall of the previous week's interactions with this set of individuals. However, participants could report contacts on weekly contact surveys by searching for contacts by name and/or email, thereby providing a "check" to see if contacts were study participants and facilitating reporting regardless of previous awareness of the study's enrollment. Weekly surveys also presented participants with two auto-populated lists of possible contacts from which they could select their previous week's contacts: one of "past social contacts" and the other, "potential social contacts." This may have biased reports towards these generated lists (possibly introducing "false" contacts), reduced reports of other contacts (that were not suggested), or possibly increased contact reporting by jogging participants' memories. Notably, the impact of nonresponse or low participation/contact reporting by some participants on the measured network could be mitigated if other network members reported contacts with these participants (i.e., reciprocal reporting).

Based on the unique characteristics of the two data collection methods, we can identify scenarios where agreement or disagreement between the networks could occur. First, it is likely that any face-to-face contacts occur in proximity of significantly less than 3 meters. We could then expect that any weekly face-to-face contacts in the final RN dataset but not in final iEN dataset were not captured by sensors for one of four reasons: 1) contact definition differences due to iEpi sampling frequency, in that any and all face-to-face contacts between a pair of participants in a given week were too short to be captured by either participant's iEpi device (i.e., the iEpi contact definition's granularity was not fine enough to detect the contact); 2) unintentional deactivation of the iEpi app due to the device losing power; or 3) intentional deactivation or "snoozing" of the iEpi app by one or both of the participants; 4) one or both participants not carrying the device with the iEpi app on their person; or 5) inaccurate reporting of a face-to-face contact (i.e., a participant reported a contact that did not actually happen). Although we can speculate that the surveys' pre-populated lists of possible contacts may have introduced false reports, it is impossible to definitively identify which, if any, reported edges did not actually occur, and must continue on the assumption that false reports were minimal. We also cannot differentiate scenarios 2, 3, or 4 in the data and treat them collectively as participant acceptance or engagement with the iEpi app. Thus, theoretically, we expect that edges present in a Reported Network but not in that week's iEpi network were either very short or not recorded due to at least one participant's lack of acceptance and/or engagement with the sensor portion of the study. We also note that given the 5 minute sampling frequency of iEpi and the weekly survey reports, all unique weekly iEpi edges and participant reports could represent any number of individual face-to-face interactions with unknown duration across a given week. Therefore, for an edge to be present in the reported datasets but not iEpi, any and all face-to-face contacts between two participants in a given week must have occurred outside the sampling frequency of the iEpi devices or when one or both participants did not have an active iEpi app on their person.

In practice, we found that the percentage of edges in $RN_3$-$RN_{10}$ that were captured in $iEN_3$-$iEN_{10}$ decreased over the course of the analyzed study period. In week 3, 91.67% of the reported dyads were also captured by iEpi, but this fell to 70.49% in week 4, and by the final study week, only 39.24% of all reported dyads were also captured by iEpi. Previous studies have found duplication of reported contacts by sensors occurred at rates of 50% (185), 70% (184), and 85% (187), which is similar to our results, particularly early in the study period. While we do not have duration data for reported contacts, we can extrapolate from the edge agreement by iEpi recorded duration. We thus conclude that because longer cumulative duration iEpi edges were more likely to also be reported than shorter cumulative duration iEpi edges, it is likely that face-to-face contacts that were reported on weekly surveys but not captured by iEpi were more than likely missed by the app due to reduced iEpi compliance, which appears to have decreased substantially over the study period. This conclusion is also supported by the significant decrease in network size and increasing number of isolates in the iEpi networks across weeks 3-10. In week 3, iEpi recorded 819 cumulative proximity contacts compared to only 201 in week 10. Similarly, while increases in isolation were not consistent across the study period, 28% of week 9's iEpi nodes were isolates in week 9 and 38% in week 10. These levels of recorded contact and isolation could represent a true reduction in contact between iEpi participants towards the end of the winter semester, but the stability in the network size and isolates in $RN_3$-$RN_{10}$ suggests that this may not be the case. It instead likely reflects, at least in part, fatigue and lowered compliance or acceptance of iEpi. We do however note that because an iEpi record of a contact event required two active apps, we cannot examine whether this reduced iEpi participation occurred differentially across participants. Although we were unable to estimate the contributions of each reason for mis-match between report and sensor edges due to small sample sizes, we suspect that differing contact definitions did not play a larger part than iEpi compliance, unlike previous short-term comparison studies that hypothesized that different contact definitions explained up to 86% of reported but not sensed contacts (185).

Conversely, there are a number of scenarios in which a proximity contact might be recorded by iEpi but not reported on a weekly contact survey. In order for an edge to be present in an iEN but not the corresponding week's RN, the proximity contact must have either: 1) not have fit the definition of face-to-face contact; or 2) fit the definition of face-to-face contact and both participants did not report the contact on a weekly survey. In the case of different contact definitions, we can imagine proximity contacts that were not "face-to-face;" these iEpi recorded contacts could have occurred across walls or between floors, or been recorded when two participants were in the same room or area of campus. Previous studies have yield a wide range of percentages of sensor contacts that were also reported (4% (187); 15% (185); 41% (184)). We found that the percentage of iEpi edges that were also reported on a given weekly survey was fairly low but it did increase over the study period: 2.69% in week 3; 15.42% in week 10. However, we should again note that overall sensor-captured contacts precipitously dropped over the study period, and the absolute number of iEpi contacts that were also reported remained at 51 or less across the study period.

It is difficult to discern the contributions of contact definition differences, recall issues, and nonresponse that resulted in iEpi to Reported disagreement. Previous studies estimate that approximately 60% of non-report of sensor contacts were due to contact definition differences and not under-reporting, or that underreporting accounted for 30-50% of discrepancies between sensor and reported networks (185,187). However, our data did not allow for direct calculation of this percentage. Importantly, we can clearly see that iEpi contacts that were also reported on weekly surveys by at least one participant had longer cumulative weekly durations than iEpi contacts that did not have corresponding reports. This result suggests that recall played a substantial role in reporting accuracy, with more time in close proximity together over the course of a given week increasing the likelihood of survey report(s). If this is the case, we can imagine that shorter iEpi contacts that were also reported on a survey might represent “important” contacts in a given participant’s personal network, as they remained memorable despite their short duration. Alternatively, the iEpi recorded cumulative duration may not be an accurate reflection of the true cumulative duration of contact between two individuals, if one or both individuals did not have the app active during portion(s) of their time together. Recall could also be dependent on when the contact took place relative to when a participant completed the weekly survey; “distance” between reporting and contact may modify the effect of duration on reporting probability. For example, while longer duration contacts may have a high probability of report regardless of the number of days between contact and survey completion, the probability of a report of shorter contacts may increase as the number of days between contact and survey completion decreases (e.g., a participant who completes the contact survey on Saturday might be more likely to report a 5 minute face-to-face contact that occurred on the previous day (Friday) than a 5 minute face-to-face contact that occurred on the previous Monday.

Notably, the Reported networks were fairly consistent across the analyzed study period in terms of node centrality and network structure, suggesting consistent participation and engagement in the weekly surveys over time. Taken in addition to the relatively high percentages of isolates, it appears that survey fatigue was minimal but that there was a segment of the study population that did not actively participate in any weekly surveys. This group however, does not appear to have had similarly low engagement with iEpi, as shown in the aggregate degree comparisons (Figure 3.4), as participants who were isolates in the all eight weekly Reported Networks had a wide range of iEpi aggregate degree.

Previous sensor measurement studies found high acceptance of wearable sensors in the short term (187,195); however, these results do not reveal the acceptability of sensors over a longer period. Given the little we know about network stability, investigating how to best measure not only static networks, but longitudinal networks, is an important avenue of research. iEpi is the first multi-week, continuously recording social network sensor study of which we are aware. Given research on longitudinal participation and contact reporting, we expected survey fatigue or panel conditioning to reduce the numbers of reported contacts over time (169,264,265), which we did not observe, but we did not have an *a priori* hypothesis regarding long-term sensor acceptability. Although iEpi participants exhibited high engagement and acceptance in the early weeks of the study (based on the high numbers of contacts and mean degree), compliance and engagement dropped significantly over time. Given the length of the iEpi study period (up to 9 weeks for participants who picked up their device at the first deployment meeting), it is reasonable to assume that the undergraduate participants began to lose interest in the study or became forgetful about keeping the device's battery charged and the app active over the lengthy study period. However, we expect that acceptance and ease of app-based data collection may be higher now than in 2013, and will likely continue to improve given ever-improving mobile battery life, increasing use of individual wearable sensors (e.g., Fitbit©, Apple watch©) (271,272), and the overall ubiquity of in mobile devices in daily life.

This study was limited in several ways. First, the two different data collection methods' differing granularities and sampling frequency required us to aggregate the iEpi data, thus resulting in the loss of detailed contact data. We additionally reduced our datasets by restricting networks according to iEpi activation dates. However, both the aggregation and matching were necessary to maximize congruence in terms of time scale and participant data. In these analyses, we characterized iEpi edges according to cumulative weekly contact duration, but there are a wealth of other possible collapsed measures by which to weight iEpi that could provide additional insights, such as number of contacts per week, longest individual contact duration per week, or number of days of contact per week. The iEpi data was also limited by allowing participants to "snooze" their data collection, which also then impacted the contacts captured by any other iEpi participants with whom they had contact while on "snooze." "Snoozing" while important for participant privacy, likely impacted the amount of contact data iEpi collected over the study period (as did the app's battery usage). Unfortunately, the iEpi dataset did not contain information about when the app was active or "snoozed," or when the device was powered off. We also acknowledge the limitations inherent in self-reported contacts, particularly panel conditioning and recall issues, which likely impacted the density and node characteristics of the observed Reported Networks, in addition to agreement with the iEpi Networks. However, reciprocal reporting does reduce the impact of nonresponse and recall on the measured network (181). Additionally, combining contact searches with suggested possible contact lists on the weekly surveys may have improved recall and/or facilitated higher reporting of repeated contacts (although it may have also added spurious contacts to the networks). The issue of recall is especially relevant given this study's survey frequency; we suspect that the accuracy of reported contacts could be improved in future studies with more frequent surveys or real-time reporting, possibly in paper or mobile device-based diaries, although these might prove burdensome in long term studies. Also, as noted, measurement error (missing contacts) could have occurred if participants did not know each other by name. One potential solution could be to incorporate a photograph roster of participants into web-based contact surveys (while acknowledging that this might raise privacy issues). Indeed, a photo-based contact diary or survey could also serve to "jog" participants' memories and reduce recall bias in contact reports, much like the pre-populated lists of possible contacts, especially for surveys spanning longer time periods.

In the future, we plan to compare the results of infectious disease transmission simulations run on the iEpi and Reported networks (as well as a combination of the two) in order to assess the impacts that the two network measurement methods have on predicted transmission patterns. Mastrandrea and Barrat investigated this topic and found that if structural similarities were identified in direct comparisons of networks observed using different data collection methods, self-reported contact data could be systematically modified and used as a proxy for more detailed sensor data in transmission modeling (195). Although we did not identify any structural similarities found across the iEpi and Reported networks, Mastrandrea and Barrat's methods, as well as network data error studies (see examples in 17,18), provide a starting point for our investigations into network transformation and edge imputation in order to more accurately estimate transmission.

Contact diaries have long been known to under-estimate contact patterns (161,170,177) and the promise of sensor-based network measurement is alluring. However, it may be that the practicalities of each method mean that neither data collection method provides a true reflection of real-world contact patterns. Given the impossibility of externally validate either dataset (172), we must rely on comparisons and conjecture in order to distill structurally and functionally meaningful contacts from high-resolution sensor data and to impute or estimate missing contacts in sensor and survey data. However, as shown in this study, network comparisons may not result in a clear preference for one method, but do give researchers the opportunity to better understand how to improve the quality of self-reported contact data, which may be a more universally feasible data collection method. Overall, quantifying agreement between measured networks can shed light on the mechanisms that produce these differences, which can then inform future study design. For instance, given that recall may have played a part in underreporting of short contacts, researchers could prompt participants to think about their brief contacts. Similarly, if sensors appear intrusive or are less engaging over time, researchers could switch to data collection methods partway through a longitudinal network study. Generally, our results suggest that while there was some overlap and duplication of contacts across sensor and reported network measurements, the most complete network would be achieved by combining the iEpi and reported data.

## 3.5 Tables

**Table 3.1** Activation dates (2013) for devices pre-loaded with the iEpi software app to iEpi study participants (N=103). Participants were included in social networks starting with the first full study week that they carried an active iEpi device (i.e., the first week after receiving and activating their iEpi device). Devices were considered activated on the date that the device first recorded an iEpi contact event.

| iEpi device activation date | N (%)[a] | Included in networks beginning with week: |
|---|---|---|
| January 28 | 59 (57.28) | 3 |
| February 4 | 20 (19.42) | 4 |
| February 7 | 1 (0.97) | 4 |
| February 8 | 5 (4.85) | 4 |
| February 12 | 14 (13.59) | 5 |
| February 20 | 1 (0.97) | 6 |
| February 21 | 1 (0.97) | 6 |
| February 22 | 1 (0.97) | 7 |
| February 27 | 1 (0.97) | 7 |

[a] Number (Percent of all 103 consented iEpi participants)

**Table 3.2** Summary of network construction and network statistics and characteristics evaluated in the iEpi and Reported Network comparisons (2013).

| | Weekly Networks (Weeks 3-10) | | Aggregate Networks | |
|---|---|---|---|---|
| | **Reported** | **iEpi** | **Reported** | **iEpi** |
| **Nodes** | Participants of iEpi sub-study; included in networks beginning with first full study week after individual iEpi device activation | Participants of iEpi sub-study; included in networks beginning with first full study week after individual iEpi device activation | Participants of iEpi sub-study | Participants of iEpi sub-study |
| **Edge data source** | Weekly surveys | iEpi-app recorded contact events, aggregated by week | Weekly surveys, aggregated over study weeks 3-10 | iEpi weekly contacts, aggregated over study weeks 3-10 |
| **Weight** | None | Summed, total contact duration for a given week | Number of weeks of contact | Number of weeks of contact |
| **Reciprocity** | Contacts reported by both dyad members: edge is identified as "reciprocal"; contacts reported by one dyad member only: edge is "unreciprocal." | None | None | None |
| **Network counts** | Nodes, edges, isolates | Nodes, edges, isolates | Nodes, edges, isolates | Nodes, edges, isolates |
| **Global network characteristics** | Density, transitivity, clustering, degree assortativity, reciprocity | Density, transitivity, clustering, degree assortativity, mean edge weight | Density, transitivity, clustering, degree assortativity, mean edge weight | Density, transitivity, clustering, degree assortativity, mean edge weight |
| **Node assessments** | Degree, indegree, outdegree | Degree | Degree | Degree |

**Table 3.3** Pairwise distribution dependence analyses between corresponding iEpi and Reported Network distributions assessed via Kendall's tau (τ) coefficient.

| Reported Network Degree Distribution | iEpi Network Degree Distribution |
|---|---|
| Weekly network (3-10) | Weekly network (3-10) |
| Weekly network (3-10) | Weekly duration[a] restricted networks (3-10)[b] |
| Weekly network indegree (3-10) | Weekly network (3-10) |
| Weekly network indegree (3-10) | Weekly duration[a] restricted network (3-10)[b] |
| Weekly network outdegree (3-10) | Weekly network (3-10) |
| Weekly network outdegree (3-10) | Weekly duration[a] restricted network (3-10)[b] |
| Aggregate network | Aggregate network |

[a] $d = iEpi\ edge\ weight\ or\ cumulative\ weekly\ contact\ duration, in\ hours$

[b] Edges restricted to: $d > 2.5\ minutes$; $d \geq 0.25\ hours$; $d \geq 0.50\ hours$; $d \geq 1\ hour$

**Table 3.4** Pairwise distribution dependence analyses between corresponding iEpi and Reported Network distributions assessed via adapted Cohen's Kappa (κ).

| Network A | Network B |
|---|---|
| Weekly Reported Networks (3-10) | Weekly iEpi Networks (3-10) |
| Weekly Reported Network (3-10) | Weekly iEpi duration[a] restricted networks (3-10)[b] |
| Reported Aggregate network | iEpi Aggregate network |

[a] $d = iEpi\ edge\ weight\ or\ cumulative\ weekly\ contact\ duration, in\ hours$

[b] Edges restricted to: $d > 2.5\ minutes$; $d \geq 0.25\ hours$; $d \geq 0.50\ hours$; $d \geq 1\ hour$

**Table 3.5** Self-reported demographic characteristics of iEpi study participants (2013).[a]

| **Characteristic** | | **iEpi participants** |
|---|---|---|
| No. of participants | | 103 (17.5) |
| Age | | 18.7 (0.1)[d] |
| Gender | Female | 68 (66.0) |
| | Male | 35 (34.0) |
| Race | White | 62 (61.4) |
| | Non-White[b] | 39 (38.6) |
| U.S. citizen | Yes | 99 (96.1) |
| | No | 4 (3.9) |
| Employed during Winter 2013 semester | Yes | 59 (62.8) |
| | No | 35 (37.2) |
| Parental education | ≤ College graduate | (49.0) |
| | > College graduate | 51 (51.0) |
| Religion | Christian | 50 (53.8) |
| | Non-Christian | 15 (16.1) |
| | Non-religious | 28 (30.1) |

[a] Data are N (%) or mean (SE).

[b] Black, Asian, Native Hawaiian or other Pacific Islander, American Indian, Alaskan Native, Multi-racial, or other race.

**Table 3.6** Network characteristics and mean node characteristics for the Reported social networks by study week (2013).

| | **Study week** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Characteristic** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| Nodes[a] | 59 | 85 | 99 | 102 | 103 | 103 | 103 | 103 |
| Edges[a] | 24 | 61 | 76 | 72 | 75 | 76 | 82 | 79 |
| Dyad reciprocity[b,c] | 0.46 | 0.46 | 0.41 | 0.46 | 0.37 | 0.43 | 0.45 | 0.39 |
| Density | 0.014 | 0.017 | 0.016 | 0.014 | 0.014 | 0.015 | 0.016 | 0.015 |
| Transitivity | 0.47 | 0.65 | 0.60 | 0.63 | 0.68 | 0.64 | 0.61 | 0.65 |
| Degree[b,d] | 0.81 (0.89) | 1.44 (1.44) | 1.54 (1.46) | 1.41 (1.46) | 1.46 (1.56) | 1.48 (1.59) | 1.59 (1.54) | 1.53 (1.61) |
| Indegree[b,e] | 0.59 (0.72) | 1.05 (1.16) | 1.08 (1.13) | 1.03 (1.18) | 1.00 (1.20) | 1.06 (1.34) | 1.16 (1.22) | 1.07 (1.21) |
| Outdegree[b,f] | 0.59 (0.78) | 1.05 (1.28) | 1.08 (1.32) | 1.03 (1.28) | 1.00 (1.31) | 1.06 (1.34) | 1.16 (1.36) | 1.07 (1.32) |
| Isolates[a] | 25 (42.37) | 29 (34.11) | 31 (31.31) | 36 (35.29) | 38 (36.89) | 37 (35.92) | 33 (32.04) | 34 (33.01) |

[a] N or N (%)

[b] Mean (standard deviation (SD))

[c] Proportion of dyads where both members (participants) reported their face-to-face contact on a given weekly survey.

[d] Unique neighbors connected to a participant.

[e] Unique neighbors who named a participant as a contact on a given weekly survey.

[f] Unique neighbors a participant named as a contact on a given weekly survey.

**Table 3.7** Network characteristics and mean node characteristics for the iEpi sensor social networks by study week (2013).

| | **Study week** | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** |
| Nodes[a] | 59 | 85 | 99 | 102 | 103 | 103 | 103 | 103 |
| Edges[a] | 819 | 504 | 517 | 550 | 344 | 356 | 274 | 201 |
| Edge weight[b,c] | 0.64 (1.44) | 0.98 (2.71) | 1.02 (2.76) | 1.04 (2.71) | 0.80 (1.80) | 1.08 (2.76) | 1.14 (2.82) | 1.05 (2.41) |
| Density | 0.479 | 0.141 | 0.107 | 0.107 | 0.066 | 0.068 | 0.052 | 0.038 |
| Transitivity | 0.86 | 0.27 | 0.28 | 0.26 | 0.23 | 0.24 | 0.24 | 0.25 |
| Degree[c,d] | 27.76 (15.10) | 11.86 (6.28) | 10.44 (6.94) | 10.78 (6.45) | 6.68 (5.42) | 6.91 (5.38) | 5.32 (4.98) | 3.90 (4.17) |
| Isolates[a] | 0 | 3 (3.53) | 11 (11.11) | 7 (6.86) | 19 (18.45) | 13 (12.62) | 29 (28.16) | 39 (37.86) |

[a] N or N (%)

[b] Cumulative contact duration (hours per week)

[c] Mean (SD)

[d] Unique neighbors connected to a participant.

**Table 3.8** Characteristics of the aggregated iEpi and Reported social networks (2013). Each aggregate network contains all dyads present in at least one of the individual weekly networks (within each network measurement method) for study weeks 3-10. Edges in each aggregated network were weighted by the number of iEpi or Reported weekly networks a given dyad was present.

| | **Aggregated network** | |
|---|---|---|
| | **Reported** | **iEpi** |
| Nodes[a] | 103 | 103 |
| Edges[a] | 121 | 1977 |
| Edge weight[b,c] | 4.50 (2.46) | 1.80 (1.35) |
| Density | 0.023 | 0.376 |
| Transitivity | 0.68 | 0.59 |
| Degree[b,d] | 2.35 (2.00) | 38.39 (18.22) |
| Isolates[a] | 24 (23.30) | 0 |

[a] N or N (%)

[b] Mean (standard deviation (SD))

[c] Number of weekly networks of a given network collection method containing a given edge (dyad).

[d] Unique neighbors connected to a participant across the aggregated study weeks.

**Table 3.9** Numbers of edges (or dyads) in each of the weekly iEpi and Reported Networks (2013), according to their cumulative weekly contact duration (in hours). Reported network edges that were not also captured by iEpi are seen in the "Not detected" column. Matched edges and edges not detected by iEpi or not reported on weekly surveys.

| | | **iEpi Weekly Network** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | **Cumulative weekly proximity contact duration recorded by iEpi app**[a] | | | | | | |
| **Study week** | **Weekly Reported Network** | **Not detected** | **2.5 min.** | **2.5 min.-0.25 hr.** | **0.25-0.50** | **0.50-1** | **≥ 1** | **Total** |
| **3** | Not reported | NA | 32 | 102 | 284 | 341 | 38 | 797 |
| | Reported contacts (dyads) | 2 | 0 | 1 | 2 | 6 | 13 | 24 |
| **4** | Not reported | NA | 107 | 183 | 59 | 62 | 50 | 461 |
| | Reported contacts (dyads) | 18 | 1 | 1 | 3 | 7 | 31 | 61 |
| **5** | Not reported | NA | 110 | 186 | 65 | 50 | 58 | 469 |
| | Reported contacts (dyads) | 28 | 2 | 5 | 2 | 7 | 32 | 76 |
| **6** | Not reported | NA | 118 | 199 | 52 | 71 | 59 | 499 |
| | Reported contacts (dyads) | 21 | 4 | 6 | 3 | 7 | 31 | 72 |
| **7** | Not reported | NA | 76 | 119 | 46 | 31 | 35 | 307 |
| | Reported contacts (dyads) | 38 | 1 | 3 | 4 | 8 | 21 | 75 |
| **8** | Not reported | NA | 100 | 102 | 41 | 38 | 35 | 316 |
| | Reported contacts (dyads) | 36 | 2 | 4 | 4 | 4 | 26 | 76 |
| **9** | Not reported | NA | 52 | 93 | 36 | 27 | 30 | 238 |
| | Reported contacts (dyads) | 46 | 0 | 7 | 3 | 7 | 19 | 82 |
| **10** | Not reported | NA | 41 | 63 | 22 | 19 | 25 | 170 |
| | Reported contacts (dyads) | 48 | 1 | 4 | 1 | 7 | 18 | 79 |

[a] Unless otherwise noted, duration is in hours per week.

## 3.6 Figures

**Figure 3.1** Flowchart of data processing, data matching, and network construction for comparisons between the iEpi Network(s) and Reported Network(s) (2013). Note that many participants attended one of two iEpi information session during the second study week, at which time they received their iEpi activated device and were given instructions on how to use the app. These large-scale meetings contributed the majority of the contacts removed in the 'Data Matching' step, as only contacts from the first full study week after iEpi activation were included in networks.

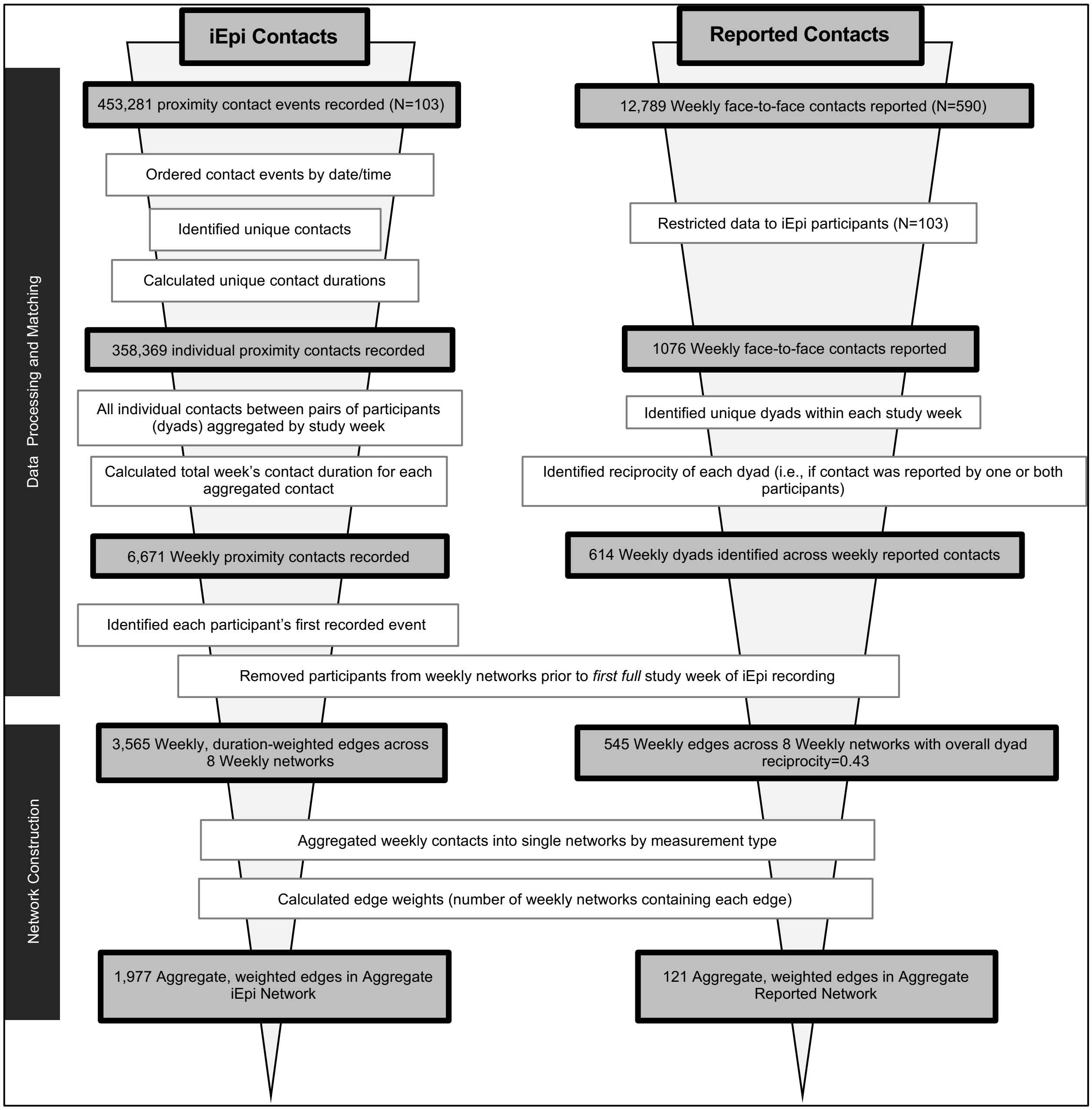


**Figure 3.2** Cumulative weekly duration for all edges in iEpi weekly networks, and overall (2013). This distribution was used when determining threshold durations for network and degree comparisons by cumulative weekly contact duration (i.e., iEpi edge weight).

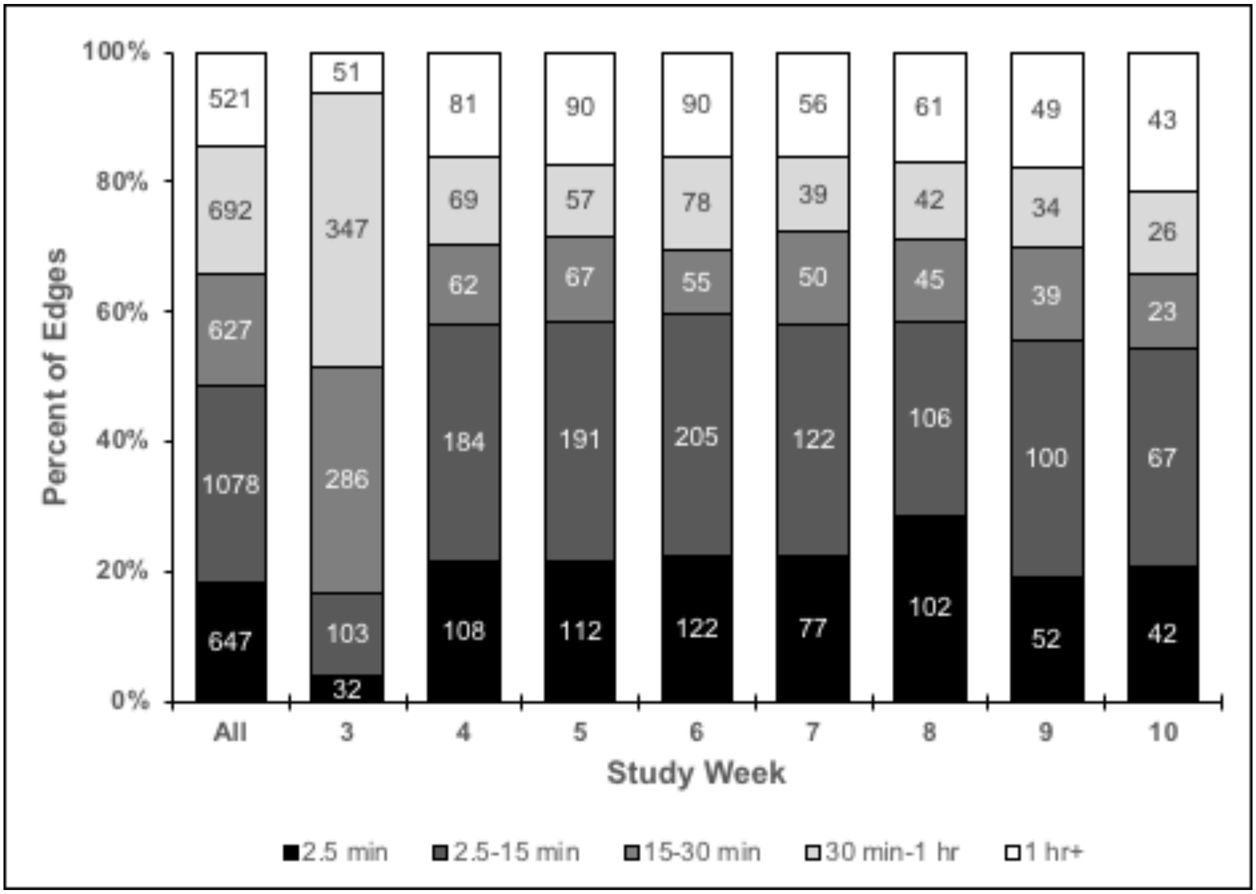


**Figure 3.3** Visualizations of the weekly (A) and aggregated (B) networks constructed from the iEpi recorded sensor data and reported weekly survey data (2013). Nodes are in identical locations in each visualization, starting with the first weekly network after individual iEpi device activation.

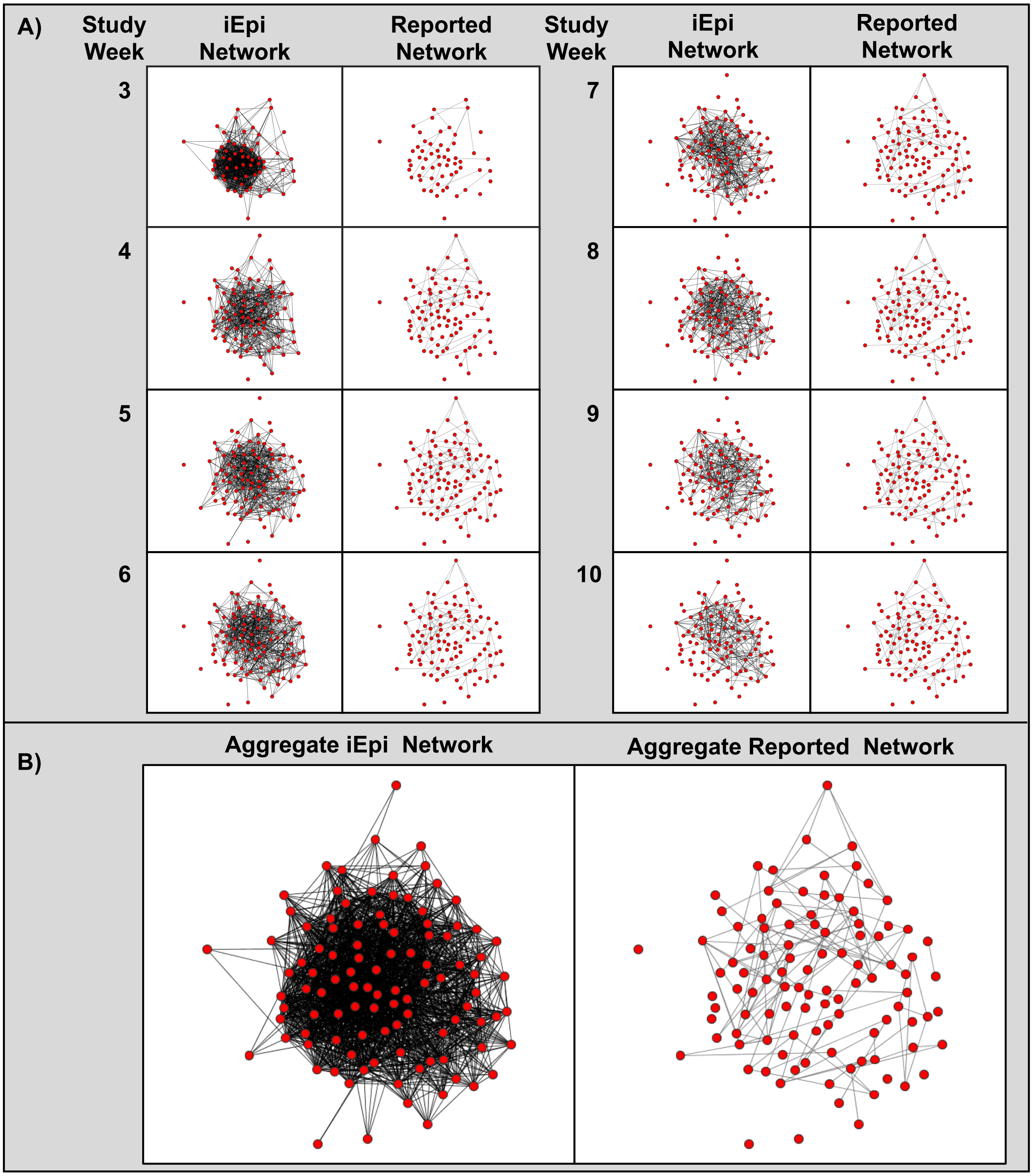


**Figure 3.4** Reported and iEpi Network characteristics across the study period and in aggregate (2013). The blue dashed line and diamonds show iEpi Network characteristics across the analyzed study period and the red line and circles show the same characteristics and time period for the Reported Networks. Panel A reflects network density over time, Panel B: transitivity, and Panel C shows the mean node degree (and standard deviation (SD)).

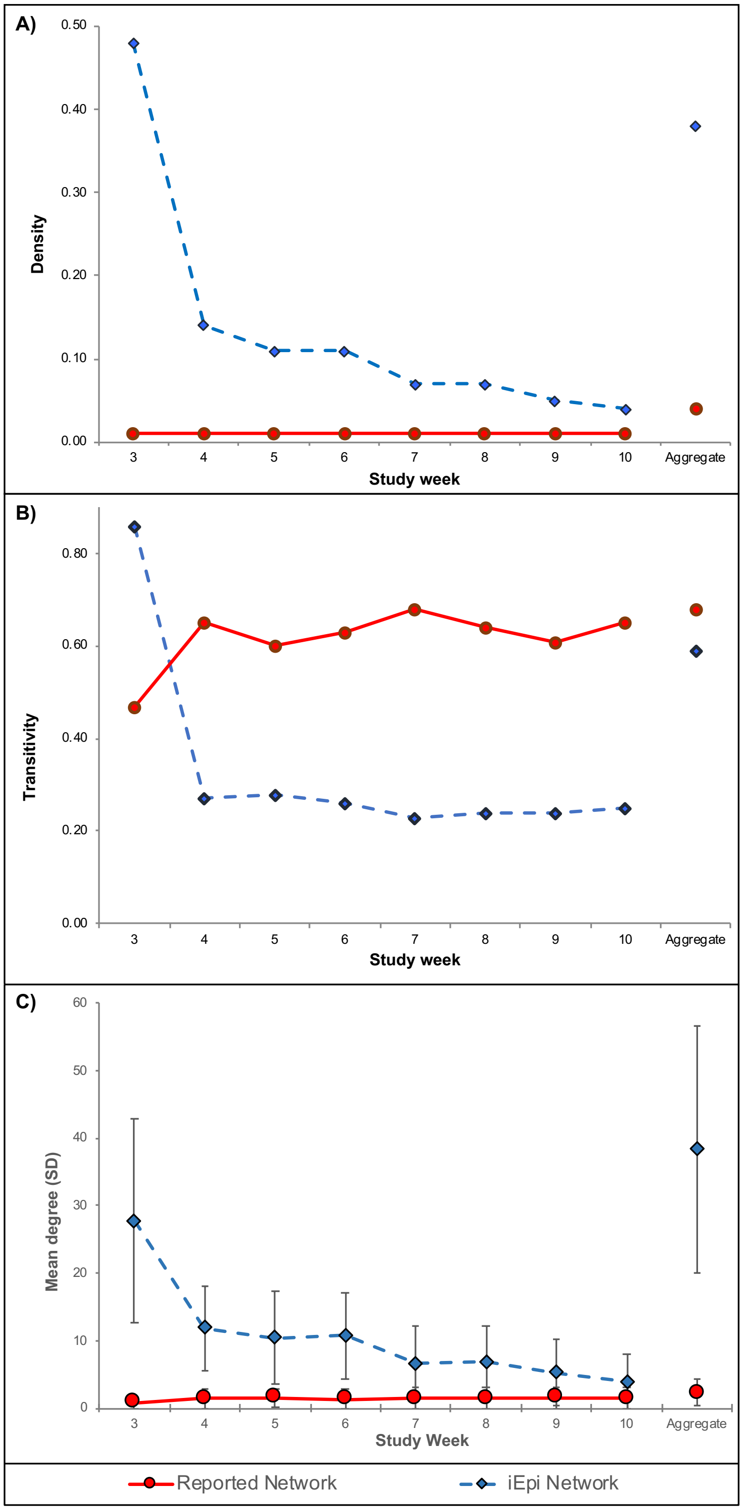


**Figure 3.5** Individual node's aggregate iEpi and aggregate Reported Network degree comparisons, with node color and size indicating the degree in the Aggregate iEpi Network (for ease of node visualization, the edges are greyed out).

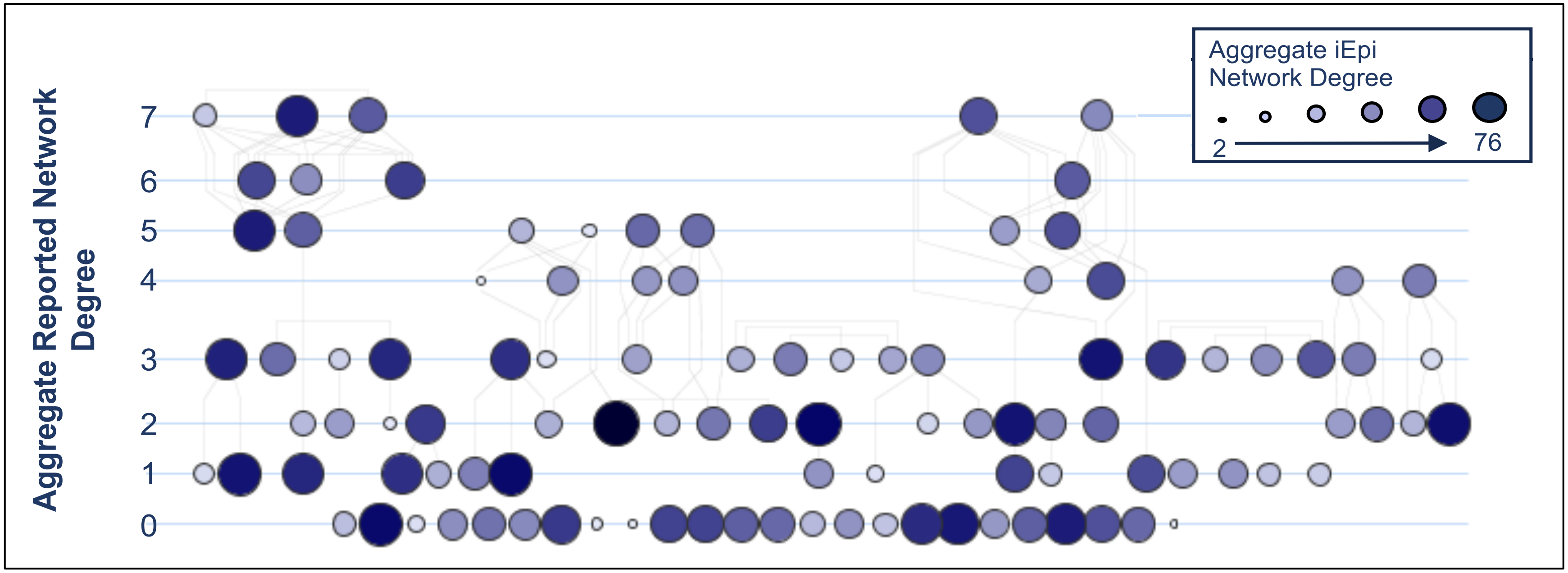


**Figure 3.6** iEpi degree distributions and Reported degree, indegree, and outdegree distributions across the study period (2013). Note that there are differing numbers of nodes in networks by week.

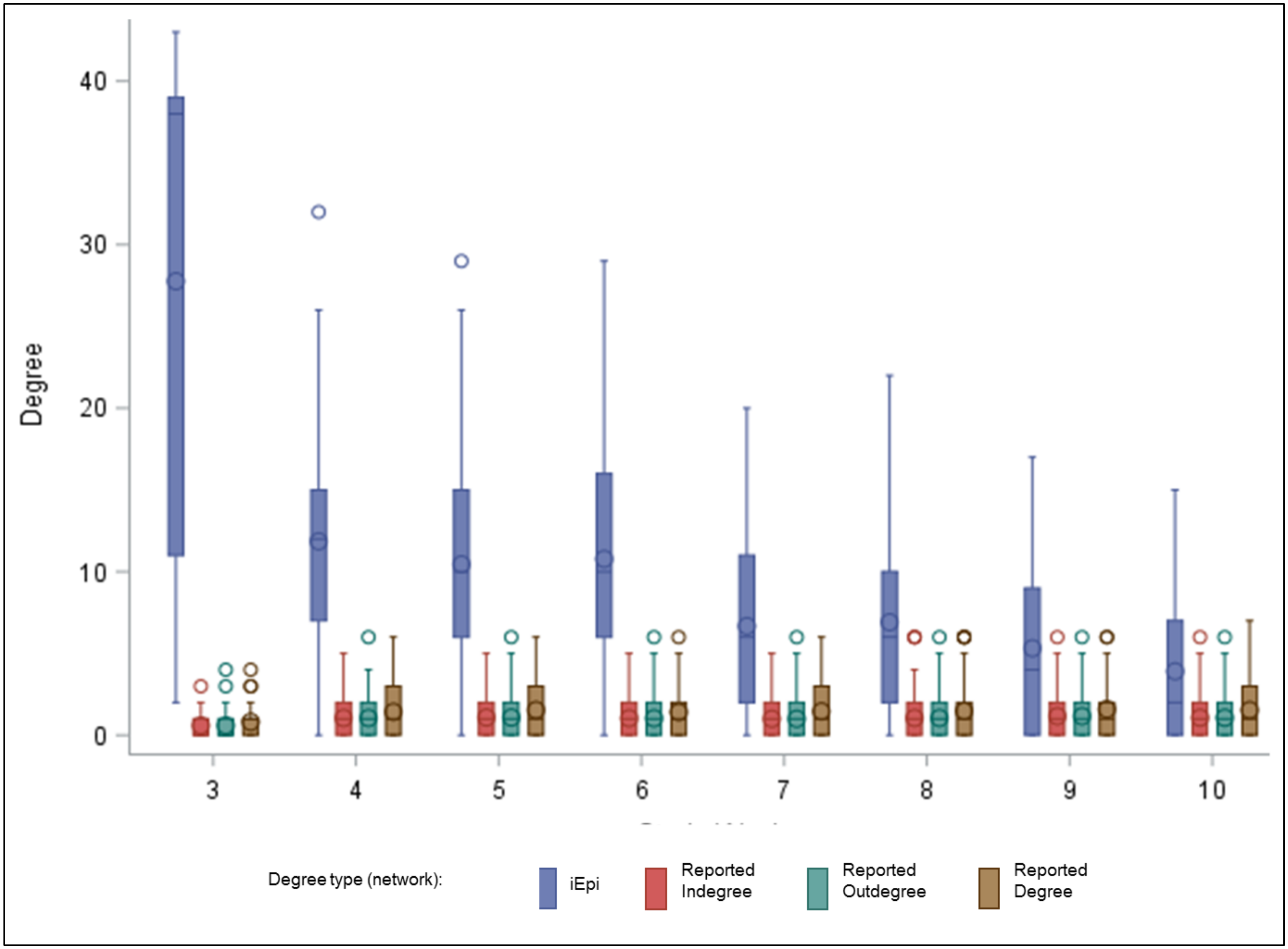


**Figure 3.7** Level of dependence or monotonic association (using Kendall's tau) between iEpi Networks degree and Reported Networks degree, indegree, and outdegree across the study period (2013). Bolded markers indicate p<0.05, representing a level of dependence significantly different from 0.

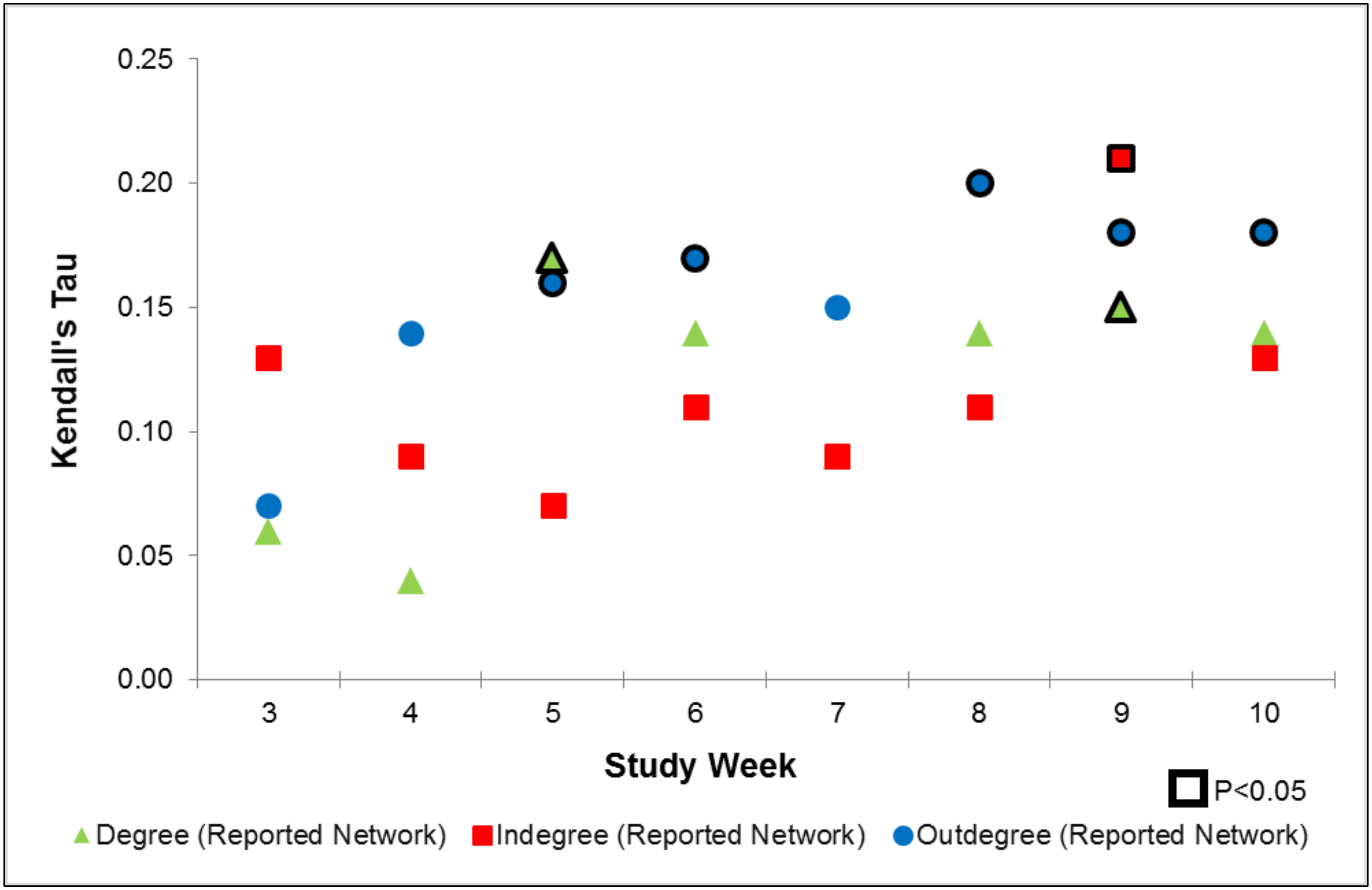


**Figure 3.8** Degree distributions for the full iEpi network and iEpi networks with edges restricted to those above the weekly cumulative duration thresholds compared to the Reported degree, indegree, and outdegree distributions across the study period (2013).

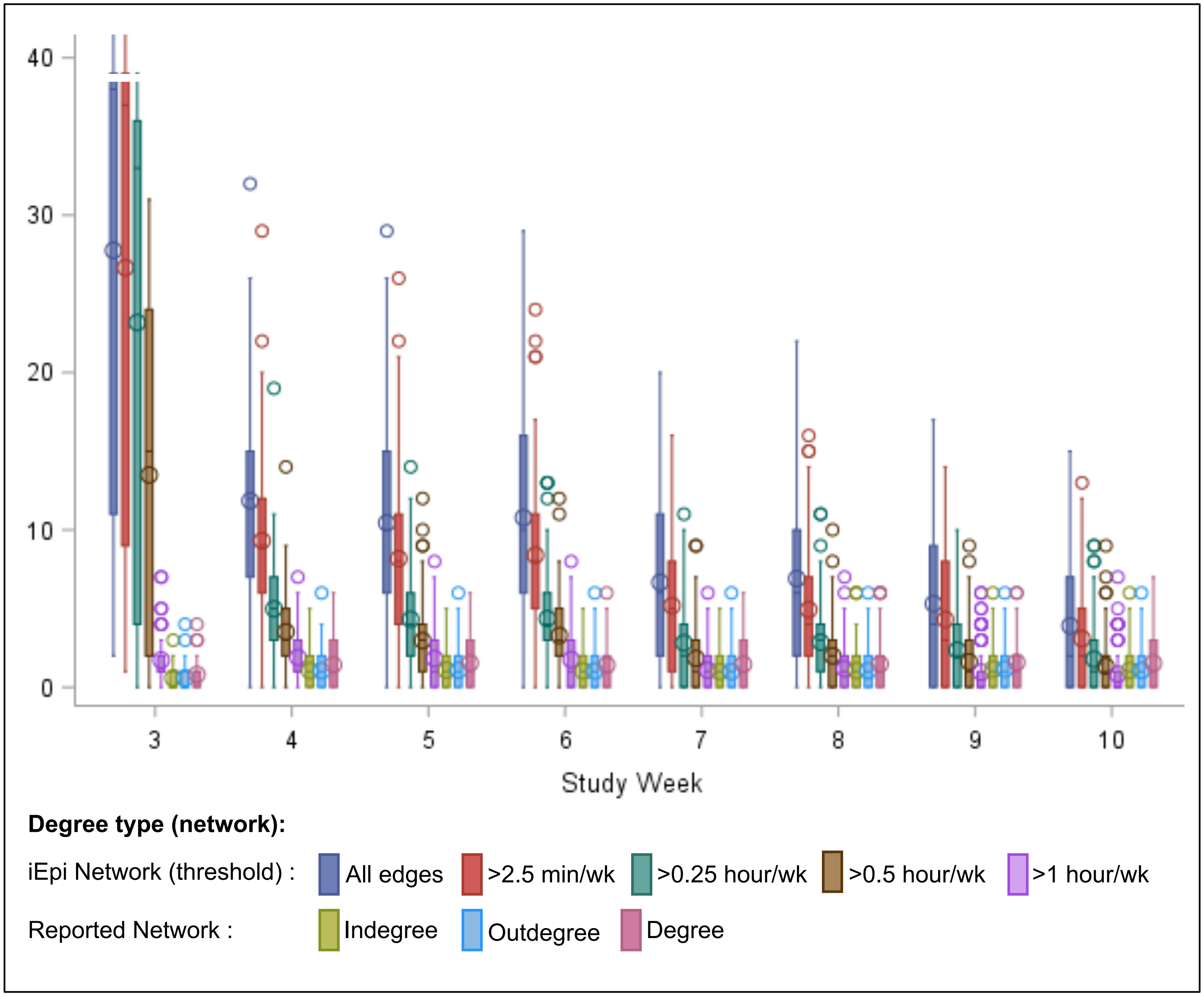


**Figure 3.9** Level of dependence or monotonic association between the degree distributions for the full iEpi network and iEpi networks with edges restricted to those above the weekly cumulative duration thresholds compared to the Reported degree, indegree, and outdegree distributions across the study period (2013). Bolded markers indicate $p<0.05$, representing a level of dependence significantly different from 0.

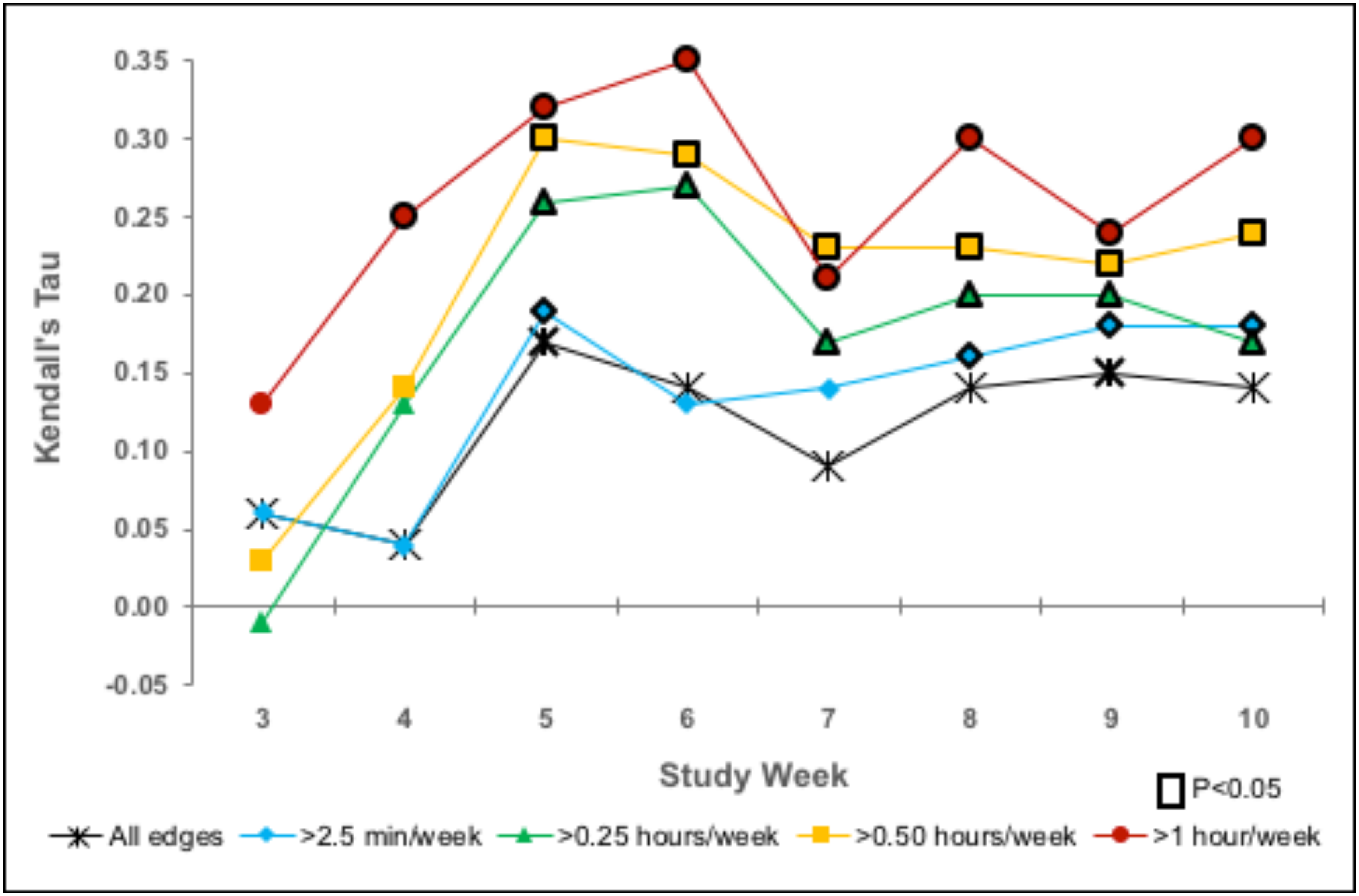


**Figure 3.10** Edges in the iEpi and Reported Networks according to whether they were captured by one or both network measurement methods (2013). The percentage of matched edges increases in the iEpi Network (left) and decreases in the Reported Network (right).

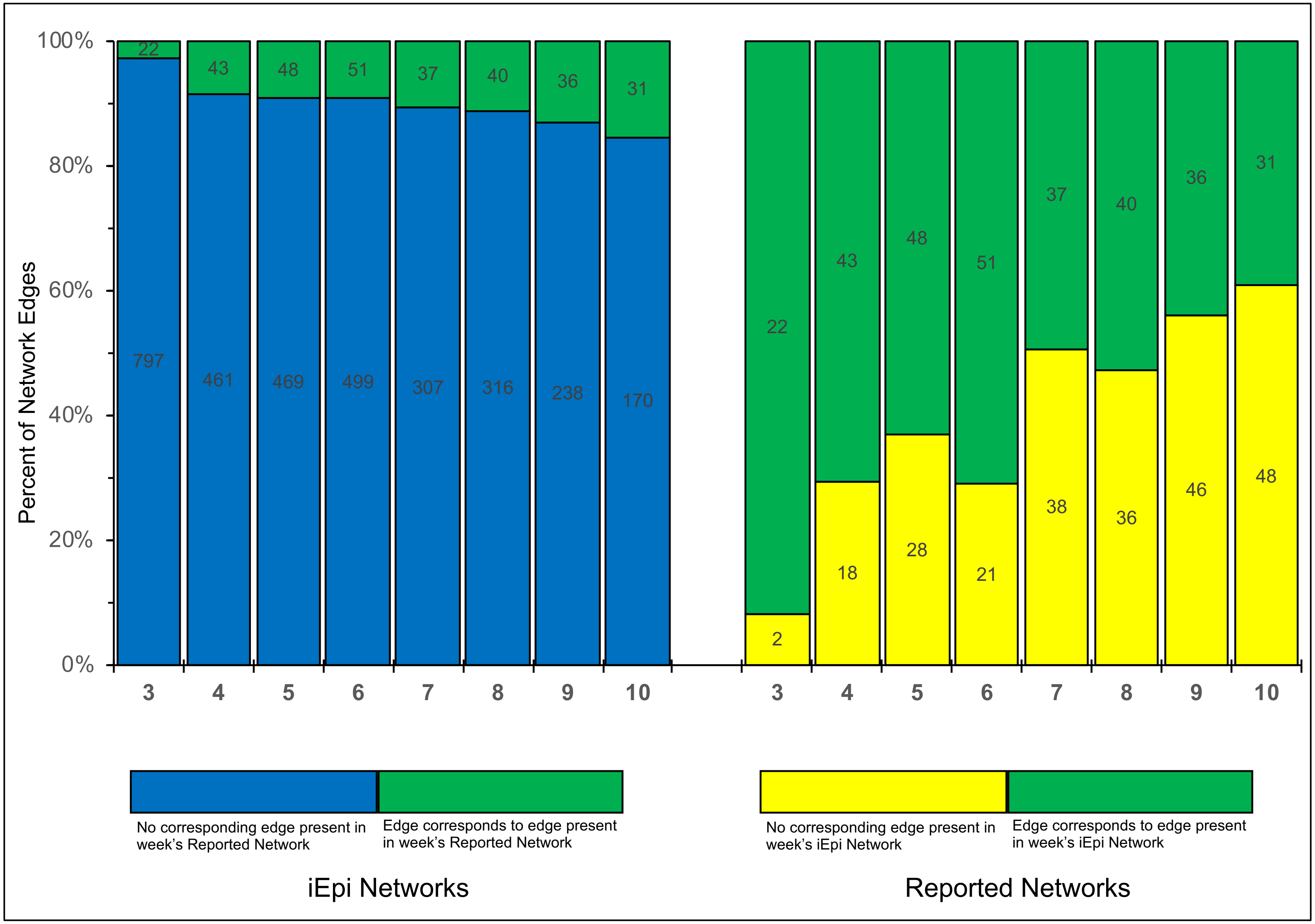


**Figure 3.11** iEpi edges' weekly cumulative contact duration by edge matched status for all edges across all study weeks (A) and broken down by study week (B) (2013). The mean cumulative weekly duration of all iEpi edges without corresponding Reported edges was 3.80 hours/week (SD: 4.40) whereas the mean duration for edges that were also reported on a weekly survey was 0.62 hours/week (SD: 1.91).

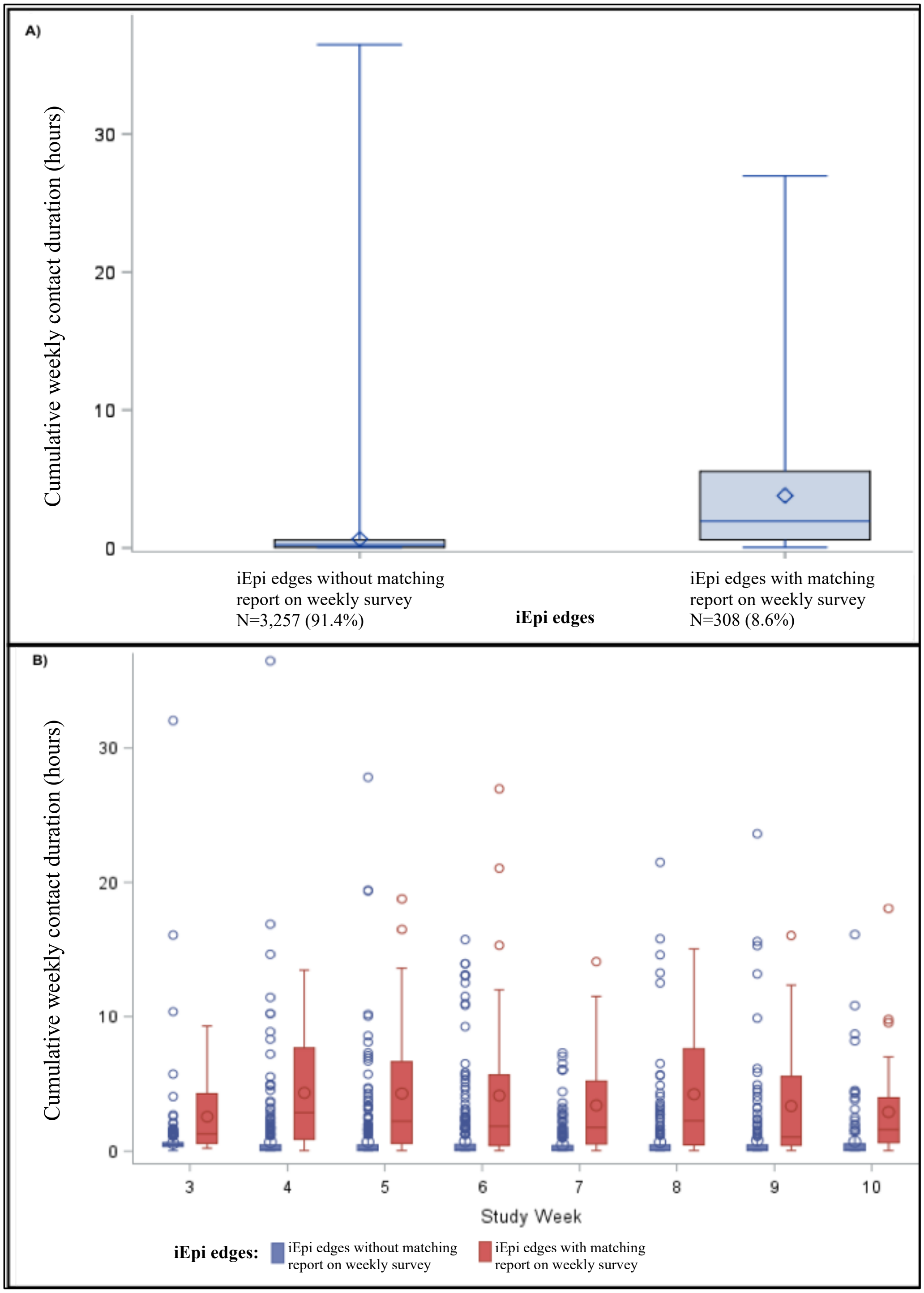


**Figure 3.12** Distributions of iEpi edges' weekly cumulative contact duration according to edge matching status across the study period (2013). iEpi edges without matching reports on a weekly survey had fairly normal duration distributions whereas the duration distribution for edge with corresponding reports on a weekly survey were skewed towards longer cumulative weekly durations.

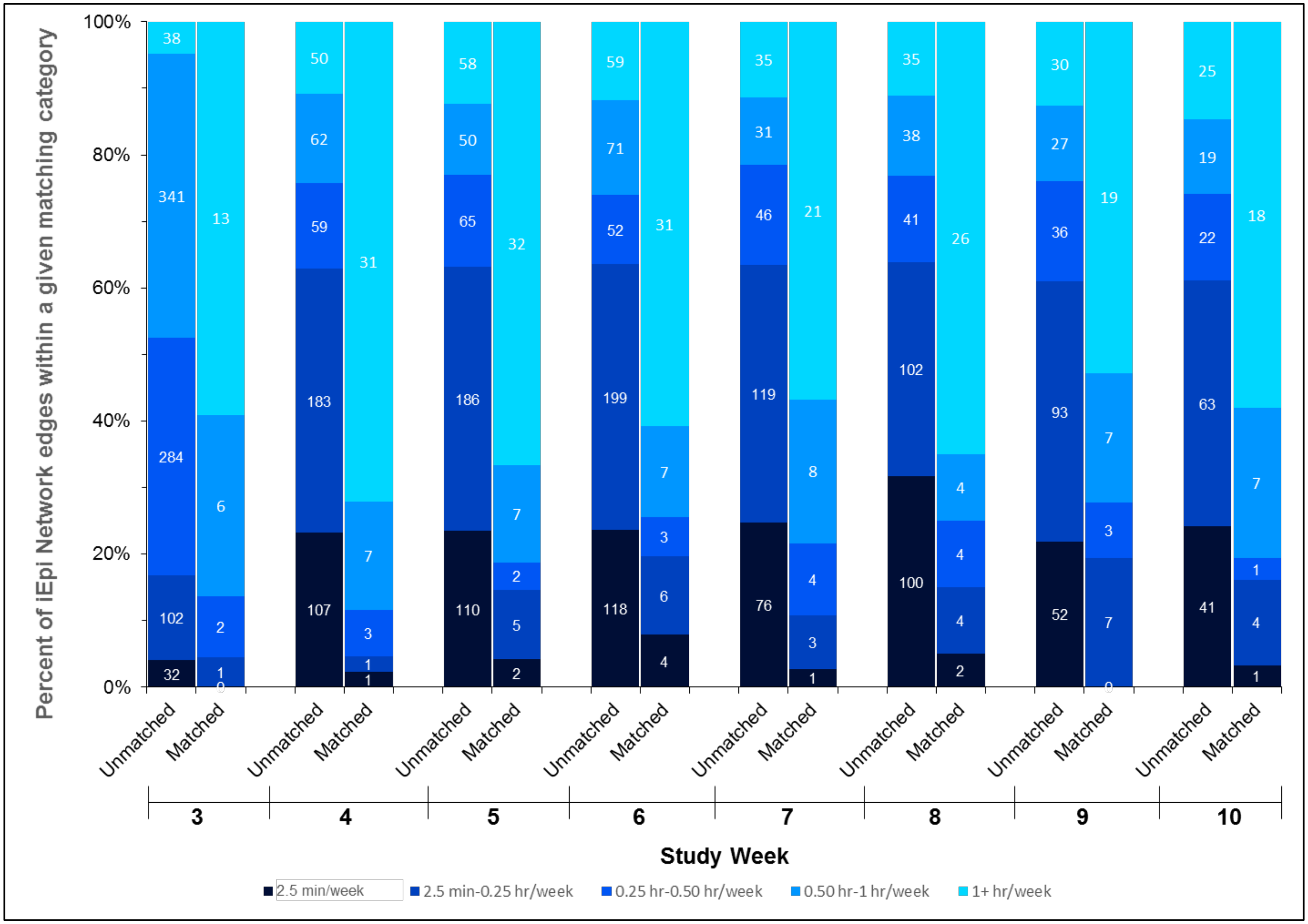
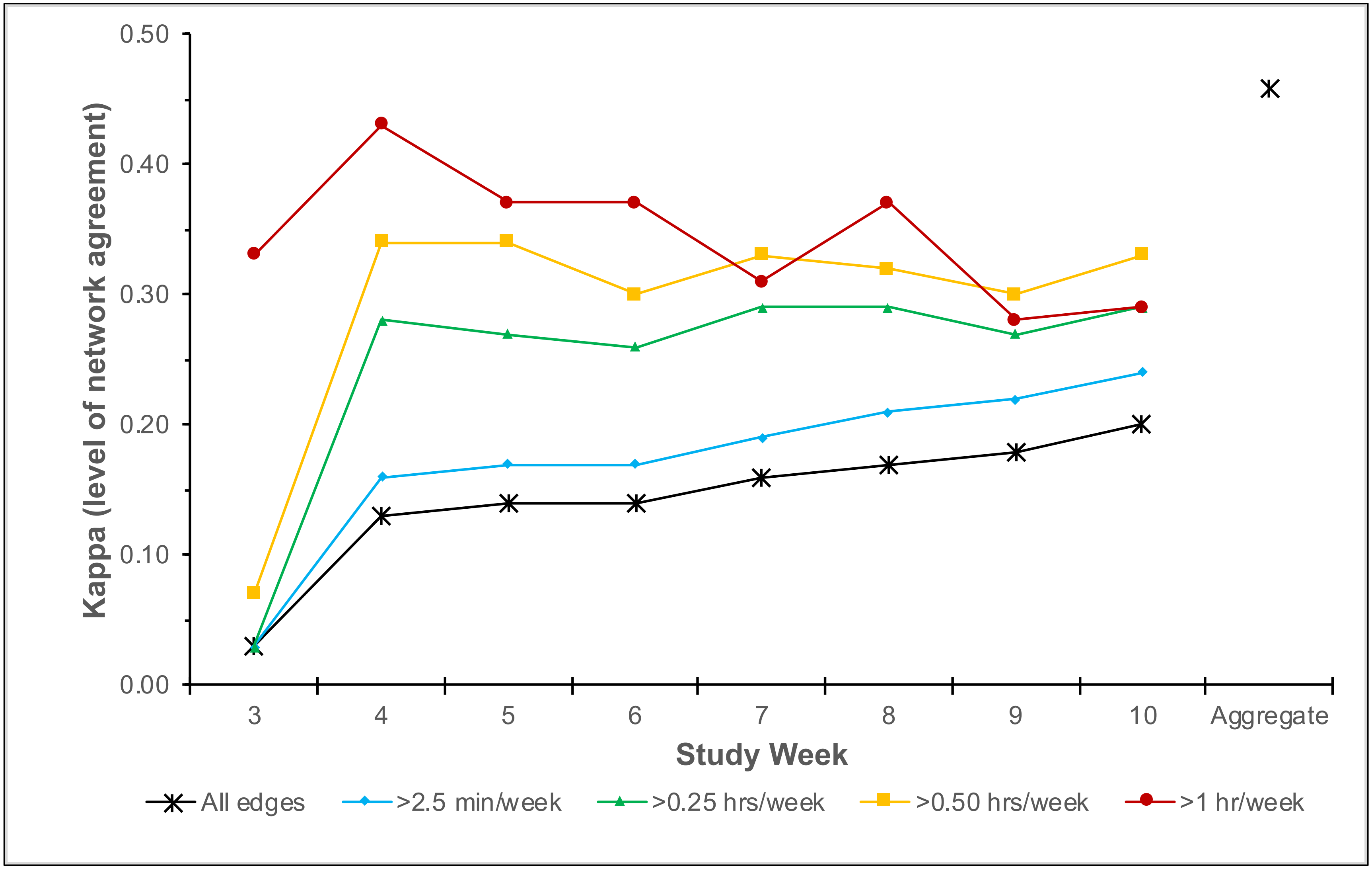


**Figure 3.13** Global network agreement between weekly and aggregate iEpi and Reported Networks over the analyzed study period (2013), as measured with the adapted Cohen's Kappa test of inter-rater agreement.



# Chapter 4

# Attribute Assortativity Sensitivity, Uncertainty, and Hypothesis Testing

## 4.1 Introduction

Social network data, sociocentric network data in particular, is by definition a set of dependent observations, with each individual in the network's presence and location enmeshed in the other network member's data; the simultaneous influence and selection processes that govern social networks invalidate traditional statistical methods (143). As a result, uncertainty, hypothesis testing, and data measurement error issues have not been fully addressed in public health social network literature, particularly for person-to-person contact networks (203). Researchers have begun to address the need for these methods by investigating the robustness (i.e., sensitivity) of network properties in the presence of measurement errors (206,209,273–276). However, few have extended these methods into uncertainty quantification and confidence interval construction, (276) and thus the reliability (i.e., the coverage of the confidence interval) of observed network characteristics is typically left un-estimated. Similarly, we have not seen any sensitivity analysis methods detailed in the literature that identify potential biases in network statistics due to participant covariate measurement error rather than measurement error in the sampled network structure. And finally, there are limited means of hypothesis testing in social network analysis (172).

Social network data consists of nodes, or individuals, interconnected by edges, or interactions or relationships. In empirical real-world networks, the network boundary, or the set of individuals in an observed social network, is defined by the enrollment approach, and the set of connections, or edges, within the defined boundary is measured with contact diaries, surveys, and/or electronic monitoring of some kind (277). The inherent dependencies within network data, in addition to the non-representative sampling (e.g., snowball sampling) typically used when measuring empirical person-to-person contact networks, impact the validity of parametric statistical methods and hypothesis testing in social network analysis (210). As such, nonparametric methods should be used in social network analysis. In particular, resampling or permutation methods are well-suited to account for non-independent data and biased group compositions (278,279). Generally, if a dataset violates the independence assumption required by traditional parametric statistical methods, resampling methods are used to systematically permutate an aspect of the observed data resulting in a "random" pattern (280) that can be compared to the observed results (281).

Measurement error in social network studies can be categorized as "traditional" survey measurement errors, including nonresponse, misclassification, and so forth, or network data measurement errors, i.e., errors in the data that describes the connections between nodes (participants). "Traditional" error and bias in observational studies has oft been addressed with sensitivity analyses that test a model's underlying assumptions to determine what impact a given variable (and its uncertainty) has on the outcome of interest (282). Thus, sensitivity analyses can estimate the robustness (or extent of bias) of a statistic in the presence of any hypothesized unmeasured confounding or effect by comparing results from the original analysis to results in which deliberate violations of the original assumptions were introduced. For example, one can in principle test the robustness of the results under the assumption of measurement accuracy by introducing known amounts of measurement error into the analyses. However, without validation data (e.g., previous studies or duplicate measurements), it is difficult to hypothesize the types and extent of errors and confounding in the data. This is especially true of social network data, where the underlying "true" network is generally unknown and may be impossible to estimate (172).

Given the likelihood that the real world network data is incomplete (202,208,209) and underestimates the "true" network's connectivity (179,180), we can assume that generally, individual degree assessments are also underestimated (179,180,206). In fact, network measurement error, while often ignored in social network analyses, is a non-trivial issue (203,205,206,276,283,284). Explorations of network data accuracy and its effects on a few network statistics (e.g., centrality) have begun but are still underdeveloped overall (205). Different types of network data error appear to have differing impacts on network statistics and characteristics, depending on the error type (206), the statistic or characteristic of interest (206), and the "true" network topology (172); it also may be impossible to separate error's sources and define its effects (209). In fact, as the "ground truth" network often cannot be known, we have no way of verifying or testing the veracity of the observed network (172) and therefore should take measures to account for possible multiple missing data effects. As such, quantifying uncertainty across a variety of data accuracy scenarios (e.g., type and prevalence of different errors) is an important step in determining how to calculate reliable estimates, as well as how much faith we should entrust in reported social network results. In turn, these determinations, in combination with *a priori* hypotheses regarding the "true" network, can (and should) inform network sampling and data collection methods (172).

Assortativity is an important social network characteristic that describes a network's homophily, or clustering according to a particular trait. Positive assortativity, or values between zero and one, indicates that contacts (edges) between individuals with similar trait(s) are more common in the network than contacts between dissimilar individuals, and disassortativity (negative assortativity) indicates preferential contact patterning between dissimilar individuals. Assortativity is particularly common to social networks (as opposed to other types of networks, e.g., animal, neural) (152,153), and describes an interaction between individual-level demographic or behavioral attributes and network topology and therefore is a particularly relevant network statistic for public health research. Importantly, Newman notes that assortative mixing can have a material effect on the robustness and behavior of networks and that models and analyses that do not account for assortative mixing can substantially misrepresent systemic effects (17). However, much like other aspects of network analysis, observed assortativity in empirical networks may be prone to bias. While observed assortativities can overestimate true assortativity due to sampling and data collection methods (152), it is unknown how other sources of measurement error can impact perceived assortativity. Given that a range of assortativities from any given sample are possible due to the variation exhibited by random networks (153,155), nonparametric randomization approaches seem a promising direction for examining error and uncertainty. However, we are unaware of studies that have applied nonparametric permutation methods to assortativity hypothesis testing to address the component causes of assortative mixing.

In order to fill these literature gaps, we have systematically investigated the observed assortativity of a large social network of undergraduate students in a secondary social network analysis of contact data collected as part of the eX-FLU study (174). The goal of this work was to: 1) develop sensitivity analysis methods to address the effects of potential measurement error in individual demographic and behavior data due to nonresponse and misclassification on observed assortativity; 2) to detail multiple resampling methods appropriate for quantifying uncertainty in observed assortativity; and 3) to quantify the probability that the observed assortativity is due to nonrandom associations (i.e., hypothesis testing), by clearly defining and implementing appropriate null models.

## 4.2 Methods

### 4.2.1 Participant self-reported covariates

We investigated network assortativity for three self-reported binary characteristics (referred to as "covariates" throughout this paper in order to distinguish them from network and node characteristics, such as degree), in the eX-FLU Year 2 study population: alcohol use, ethnicity, and gender. Baseline alcohol was reported on the first weekly survey of the study period. Participants reported their gender and ethnicity during enrollment, which was ongoing from the previous fall semester (2012) and through the start of the study period in the winter 2013 semester. The survey questions from which this data was abstracted can be found in Appendix A.2. Participants' alcohol use was dichotomized into baseline drinkers and non-drinkers based on their self-reported current alcohol use (see Chapter  2.2 for details), male or female, as reported on the enrollment survey, and as self-identifying their ethnicity as either Hispanic or Latino or non-Hispanic or Latino. These three variables were selected in order to capture: 1) a range of nonresponse levels in the dataset (alcohol use: 28%, ethnicity: 10%, and gender: 5%) and 2) both expected differential and non-differential misclassification, with respect to these covariates (differential: alcohol use, non-differential: gender and ethnicity).

### 4.2.2 Network construction

To avoid complicating our assortativity analyses with multi-edges, we constructed a simple digraph (directed) graph using data from the first weekly contact survey of eX-FLU Year 2's study period only. All Year 2 enrolled participants were included in the network as nodes and edges were drawn between participants who reported face-to-face contact with each other on this first weekly contact survey. Edges were directional, emanating from the participant who reported the contact and directed towards the participant with whom the contact was reported. This allowed for a maximum of 2 edges between a pair of participants (i.e., a dyad), given that both participants reported the contact (i.e., reciprocal edges). If the contact was only reported by one dyad member, the single edge within the dyad was unreciprocal.

### 4.2.3 Network and node characteristic analyses

We counted the numbers of nodes, edges, and isolates (i.e., nodes without any contacts) in the social network and measured the network's density, transitivity, and global dyad reciprocity. We also assessed individual node-level centrality measures and calculated their network means, including: degree, indegree, outdegree, and node-level dyad reciprocity (see Chapter 1.2.2 for definitions of these measures). The network was visualized using Visone 2.17 (Konstanz, Germany).

For each of the three identified participant self-reported binary covariates (alcohol use, ethnicity, and gender), we assessed the distributions of the node characteristics (degree, indegree, outdegree, and node-level dyad reciprocity across both binary choices; i.e., drinker/non-drinker, male/female, and Hispanic/non-Hispanic), as well as for participants missing data for each covariate. Statistical differences in the distributions between the two groups with data (e.g., male and female), as well as differences between participants with and without data for each covariate were tested using bootstrapped resampled differences in means. In each resample, the distribution of each network characteristic was sampled with replacement, using the original sample sizes, and a new difference in means was calculated. The P-value represents the proportion of resampled differences in means that were as or more extreme than the difference in means assessed from the original study population without resampling.

### 4.2.4 Observed attribute assortativity

Alcohol use, ethnicity, and gender assortativities for were assessed for the measured social network (see Chapter 1.2.5 for detailed information on assortativity), and for each attribute assortativity, sensitivity analyses, uncertainty quantifications, and hypothesis testing were conducted. Note that we refer to assortativity as “attribute assortativity” throughout this paper in order to distinguish it from degree assortativity, a commonly calculated network statistic describing preferential attachment according to degree (57). Network visualizations highlighting attribute assortativities for each covariate were created using Visone 2.17 (Konstanz, Germany).

### 4.2.5 Attribute Assortativity sensitivity analyses

In order to assess the potential impacts that nonresponse or differential and non-differential misclassification of individual-level self-reported demographic or behavioral data could have on network attribute-based assortativity, we developed and conducted two types of sensitivity analyses using randomization methods.

*Nonresponse sensitivity analysis*

Given that attribute assortativity does not account for missingness in node data, it is highly likely that nonresponse levels will measurably impact a network’s observed attribute assortativity. To assess the impact of participant non-response, or missing data, on observed assortativities in the social network, we designed and conducted sensitivity analyses for alcohol use, ethnicity, and gender assortativity. Generally, the goal of these analyses was to compare the observed attribute assortativity to the mean attribute assortativity that would have occurred in a network without any missing data for a given covariate, across the range of potential distributions of said covariate. For each covariate, we identified participants with and without missing data. First, all participants with missing data were assigned a single value for the given covariate, maintaining the proportion of a given covariate for the total study population, and a new observed attribute assortativity was calculated and recorded for this network with complete covariate data. Next, we systematically randomized possible distributions of the covariate across the population segment with missing data, and each distribution was randomly assigned 10,000 times. For each randomization, a new attribute assortativity for the network without any missing data was calculated and recorded. The mean and standard deviation of generated attribute assortativities for each distribution of the covariate was plotted.

For example, 58 participants, or 9.83% of the population, was missing ethnicity data, 27 (4.58%) identified as Hispanic or Latino, and 505 (85.59%) self-identified as neither Hispanic nor Latino.

We first assigned all 9.83% of participants without ethnicity data as being of Hispanic origin, thus maintaining the segment of the non-Hispanic population at 86% and raising the Hispanic percentage to 14.41%. Attribute assortativity for this new non-missing data network was recorded and plotted. Next, we increased the overall percentage of non-Hispanic participants by small increments by randomly selecting nodes with missing data to identify as Hispanic (assigning the remaining missing data nodes a non-Hispanic ethnicity). The attribute assortativity was assessed and recorded for each increment. The procedure was repeated for each increment 10,000 times, and the mean and standard deviation of those randomizations was plotted.

### *Misclassification sensitivity analysis*

We investigated the potential impacts of differential and non-differential misclassification due to under- or mis-identified alcohol use, ethnicity, and/or gender on the network's observed assortativities using randomization-based sensitivity analyses similar to those described above for potential nonresponse bias. We first hypothesized whether we expected misclassification of our binary variables would likely be differential or not. For ethnicity and gender, we hypothesized that levels of mis-reporting gender and ethnicity would be non-differential, e.g., men and women would have similar levels of misclassification. For alcohol use, we hypothesized that drinking under-reporting would most likely be confined to non-drinkers, or differential with regards to alcohol use. We additionally hypothesized that participants under the legal drinking age of 21 might be more likely to underreport their alcohol use than self-identified non-drinkers ages 21 and over, so we conducted two analyses for alcohol use: in the first, we assessed the potential effects of misclassification within all non-drinkers, and in the second, the potential effects of misclassification within underage non-drinkers.

### *Non-differential misclassification sensitivity analysis*

For variables with hypothesized non-differential misclassification, we performed the following randomization procedures for the sensitivity analysis:

1. While keeping the list of nodes with missing data constant, we randomly changed or reversed responses (classifications), across a specified percentage of nodes in each of the two groups (e.g., randomly changing the gender of 5% of self-reported male participants to female and 5% of self-reported female participants to male) and reassessed the

attribute assortativity after each of the 10,000 randomizations.

2. For each randomization, we recorded the new attribute assortativity and plotted the mean and standard deviation of observed assortativities for each new, randomly generated distribution of the variable of interest. We repeated each of the 10,000 randomizations for a range of misclassification percentages, i.e., we reassigned gender to a random sample of 5% of male and 5% of female participants, then 10%, 15%, and so forth, recording the mean attribute assortativity of 10,000 randomizations for each potential level of misclassification.

*Differential misclassification sensitivity analysis*

For variables with hypothesized differential misclassification (alcohol use), we performed the following randomization procedures for the sensitivity analysis:

While keeping the list of nodes with missing data constant, we randomly changed or reversed responses (classifications), across a specified percentage of nodes in the group with hypothesized misclassification (e.g., randomly changing 5% of self-reported non-drinkers to drinkers) and reassessed the attribute assortativity after each of the 10,000 randomizations.

For each randomization, we recorded the new attribute assortativity and plotted the mean and standard deviation of observed assortativities for each new, randomly generated distribution of the variable of interest. We repeated each analysis for a range of non-differential misclassification percentages.

### 4.2.6 Attribute assortativity uncertainty

In addition to assessing the direction and magnitude of potential bias in attribute assortativity calculations resulting from individual data measurement error, we also investigated and quantified the effect of network data sampling on attribute assortativity. We calculated estimated endpoints of 95% confidence limits of the assortativities using bootstrapping and jackknifing (285,286). These methods sample or delete random nodes and/or edges, modeling the sampling process by resampling the original network to generate a new ‘sampled’ subnetwork. We tested these methods across a range of levels of random node and edge deletions. The goal of these network permutations was to determine the range of likely values of attribute assortativity for the covariate of interest in the sampled social network, as well as the range of confidence intervals that result from different resampling procedures. In so doing, we addressed the robustness of attribute assortativity given two types of potential network measurement errors, specifically node or edge deletion (206). We analyzed these two types of error separately and under the assumption that errors occurred randomly on the network, acknowledging that in practice: 1) there are likely a number of interacting influences that effect network measurement error, and 2) real world network data likely contains mix of multiple error types (172).

*Jackknifed models*

The jackknife (or deletion) models evaluate the uncertainty in the network data under the assumption that a specified proportion of the "true network's" edges or nodes (selected at random) were not observed, and were therefore missing from the dataset. To quantify the uncertainty in observed attribute assortativity, we conducted jackknifed random removal of edges and nodes across a range of removal amounts (a single edge/node, 1%, and the range of 5-100% at 5% intervals), on the eX-FLU week 1 simple digraph. We included the jackknife removal of a single edge or node based on the Newman's recommendation for uncertainty calculation using this approach (57). However, given that for medium to large networks (e.g., such as the 590-node network considered here), systematic, jackknifed removal of any single node or edge would likely produce a negligible 95% confidence interval, we included the range of percentage-based jackknife in our analyses. We calculated two sets of 95% confidence intervals for alcohol use, ethnicity, and gender assortativities (one set using edges as the removed structure and one set using nodes) using following process, repeated 10,000 times:

1. Randomly remove X amount of edges or nodes from the observed social network (1 edge/node, 1%, and the range of 5-100% at 5% intervals).
2. Calculate and record this reduced network's attribute assortativities for all covariates of interest.

This process yielded empirical distributions of 10,000 assortativities by each of the given covariates. The endpoints of the 95% confidence interval (CI) for a given attribute assortativity was then defined as the two 2.5% tails of the empirical distribution.

*Bootstrapped models*

In theory, bootstrapped models model the original network sampling by resampling the observed data in order to assess the level of error, or uncertainty, in the dataset. We calculated attribute assortativity 95% confidence intervals for the three selected covariates using bootstrapped edge resampling on the network 10,000 times. For each resampling, a new list of directed edges with size equal to the original number of edges, was randomly sampled with replacement from the network's list of edge and added to an empty network containing all nodes in the observed network. However, it is important to note that given a simple digraph with a maximum of two edges (one in each direction) per dyad, as is the case here, or a simple undirected graph, this process results in a in the resampled graph with fewer edges than observed in the original graph, as any sampled edge will only be added to the new network once, regardless of how many times it was sampled. This therefore means that this bootstrapped edge resampling process is in effect the same as the jackknife edge removal process described above, but with a random edge removal amount.

### 4.2.6 Attribute assortativity hypothesis testing

In addition to quantifying attribute assortativity uncertainty and robustness given probable individual and network measurement error(s), we conducted hypothesis testing for attribute assortativity using null models. We hypothesized that the observed assortativities for alcohol use, ethnicity, and gender in the observed social network had not arisen by chance and were a result of preferential face-to-face contact between similar individuals. In order to test these hypotheses, we isolated and quantified the effects of node covariates on network structure above random chance using null models. We defined two distinct null models and performed randomization or permutation tests using on our observed network and these null models, with the goal of quantifying the range of attribute assortativity values (for each covariate) that could have arisen by chance. We implemented each of the two null models by generating "random" networks based on the models, in an attempt to isolate the hypothetical effects of the covariate(s) on edge formation. In other words, we simulated a range of networks that could have occurred "by chance," given the characteristics of the observed network's set of nodes or edges. We defined our null models to control for either: 1) observed distributions of participant covariates or, 2) observed network edges.

*Node null models*

Given the hypothesis that a single node covariate was not associated with edge formation in the observed network, we implemented a null model isolating the effects of a single node covariate, e.g., gender, on edge formation. This null model was defined as a network with randomly allocated values of node covariate of interest (maintaining the overall distribution of the variable, including missingness), but identical to the observed network in all other aspects, including dyads, and directed edges, and all other node covariates other than the variable of interest.

We conducted the following process 10,000 times for each covariate of interest in order to draw distributions of assortativities that could occur by chance, i.e., from the null model, and calculate statistical significance levels of our observed attribute assortativities:

1. Randomly distribute the observed distribution of the specified covariate (e.g., gender) across the nodes in the observed network, maintaining the overall distribution of values (e.g., randomly allocate a gender to each node, keeping the total number of male and female participants constant across iterations. As nodes with missing data for a given covariate are not included in attribute assortativity calculations by definition, we incorporated covariate nonresponse levels into our randomization scheme in an attempt to account for any confounding by nonresponse. We treated missingness as a unique value of the covariate, and randomly allocated the covariate's distribution, including missing values, across all nodes.
2. Calculate and record the randomly generated network's attribute assortativity.

We then plotted the histogram of attribute assortativity frequencies from the randomized networks, and calculated the attribute assortativity P-value as the proportion of the distribution of randomly generated network assortativities that were as or more extreme than the observed network's attribute assortativity. Using a statistical significance level of 0.05, we rejected the null model if more than 5% of the randomized networks exhibited attribute assortativities as or more extreme than the observed network, we would then reject the null hypothesis of non-preferential attachment with respect to the covariate of interest, i.e., an attribute assortativity of 0.

### *Edge null models*

Given the same hypothesis as described above in the node null model hypothesis test, that covariates were not associated with edge formation in the observed network, we implemented an edge-based null model to isolate the effects of covariates, e.g., a node's gender, on edge formation. This null model was defined as a network containing the observed network's nodes and same density as the observed network, but a randomly generated set of edges drawn from the list of possible edges for the observed network, i.e., the complete list of combinations of network nodes. In other words, these randomly generated networks had identical nodes (keeping observed covariates constant) and density to the observed network, but randomly, not preferentially, drawn edges.

We conducted the following process 10,000 times in order to draw distributions of attribute assortativities that could occur by chance, i.e., from the null model, and calculate statistical significance levels of our observed assortativities:

1. Generate a network containing the observed network nodes and no edges. Next, randomly select the same number of edges as the observed network from a list of possible network edges, or dyads, equivalent to all possible combinations or permutations of network nodes. For a directed network, as ours was, this list contains a total of $|V| \cdot |V-1|$ possible edges from which to draw, where $V$ is the number of nodes in the network. For undirected networks, this list would be equivalent to all possible permutations of the network's nodes, of size $\frac{(|V| \cdot |V-1|)}{2}$. Note that given that the edge-based null model construction maintains the observed set of nodes, this null model and hypothesis, in effect, controls for the *combination* of node attributes on edge formation, as opposed to individual attributes. Therefore, we can calculate and record all assortativities of interest for each generated edge-based null model-network.
2. Calculate and record the randomly generated network's assortativities.

For each covariate, we plotted the histogram of attribute assortativity frequencies from the randomized networks, and calculated the attribute assortativity P-value as the proportion of the distribution of randomly generated network assortativities that were as or more extreme than the observed network's attribute assortativity. Using a statistical significance level of 0.05, we rejected the null model if more than 5% of the randomized networks exhibited assortativities as or more extreme than the observed network, we would then reject the null hypothesis of non-preferential attachment with respect to the covariate of interest, i.e., an assortativity of 0.

## 4.3 Results

### 4.3.1 Network and node characteristics

Network structure and characteristics of the Year 2 Week 1 simple digraph are shown in Table 4.1 and the network is shown in Figure 4.1. The network contained all 590 consented participants in the eX-FLU study (Year 2). There were a total of 1275 directed edges between them, with a global dyad reciprocity of 0.29, i.e., 29% of the contacts in the network were reported by both members of a given dyad (as opposed to just one dyad member). The network density was low, at 0.023, in part due to the 122 (21% of nodes) isolated, but the network exhibited high clustering, with a transitivity equal to 0.68. The mean degree (i.e., number of unique neighbors or contacts) was 3.50 (SD: 3.97), and the mean in- and out-degrees were both 2.16 (SD: 2.57 (indegree); 3.38 (outdegree)).

Table 4.2 shows the distribution of network characteristics by alcohol use. The distributions of all node centrality measures were significantly different across these three baseline alcohol use responses (drinker, non-drinker, nonresponse). Drinkers, comprising 26% of the study population (45% non-drinkers, 28% missing data) were the most central network members, on average, with significantly higher degree, indegree, and outdegree than non-drinkers and missing data participants. Missing alcohol data participants had the lowest mean centralities for all seven characteristics.

Participants who self-identified as Hispanic (5%) were similarly central in the network across all analyzed network characteristics (see Table 4.3). Participants without ethnicity data were less central than those with ethnicity data, with lower mean degree, indegree, outdegree, and dyad reciprocity ($P<0.05$ for degree, outdegree, and dyad reciprocity).

The only significant gender differences in centrality were found between those with and without gender data, with those missing gender data having lower centralities all analyzed measures (see Table 4.4; $P<0.05$ for degree, indegree, outdegree, and dyad reciprocity).

### 4.3.2 Observed attribute assortativity

The social network was fairly assortative across all three analyzed covariates (attributes). Generally, assortativity absolute values greater than approximately 0.10 are accepted as “high” or relevant to network structure (17,173) and the observed attribute assortativities in the eX-FLU social network were 0.19 (alcohol use), 0.18 (ethnicity), and 0.17 (gender). Network visualizations illustrating attribute assortativity within nodes with data and nodes with missing data for each of the three analyzed covariates are shown in Figures 4.2–4.4.

### 4.3.3 Attribute assortativity sensitivity analyses

*Nonresponse sensitivity analysis*

The sensitivity analysis assessing the potential impacts of the 28% alcohol use nonresponse on the observed alcohol use shows that the impact of missing data on the observed attribute assortativity was minimal (Figure 4.5). When various levels of alcohol use were repeatedly randomized to the missing data segment of the study population, the generated network assortativities are lower than the observed value, but remain high, dropping from the observed value of 0.19 to a range of mean values of approximately 0.17 to 0.15, with narrow variation ranges across the generated resampled proportion of drinkers. The largest change in attribute assortativity was observed when the generated network with no missing data had approximately 53% drinkers; in this case, the attribute assortativity was 23% lower than the observed value.

Observed ethnicity assortativity, which had approximately 10% nonresponse, or missing data, appears to have been somewhat impacted by this missing data (Figure 4.6). When all nodes with missing ethnicity data were assigned as “Hispanic,” the ethnicity assortativity dropped from the observed value by approximately 38%, from 0.18 to 0.11. As we incrementally decreased the proportion of missing data nodes randomly identified as Hispanic, the ethnicity assortativity increased linearly to a maximum of 0.15, or approximately 87% of the observed ethnicity assortativity. Across the range of assigned ethnicities tested, the network without missing node ethnicity data exhibited positive, measurable assortativity by ethnicity, suggesting that although the observed ethnicity assortativity on the measured network was likely an overestimation of preferential clustering by ethnicity, the underlying ethnicity assortativity was still positive and relevant to the network structure.

For gender, which had a very low nonresponse rate (5%), the sensitivity analysis shows a negligible impact of missing data on the observed gender assortativity (Figure 4.7). In these results we see a very narrow range of gender assortativities within and across the generated networks with increasing female proportions.

*Misclassification sensitivity analysis*

Our non-differential misclassification sensitivity analysis results are shown in Figure 4.8 (A: ethnicity, B: gender). Across the range of non-differential misclassification levels, both ethnicity and gender assortativities exhibit parabolic trends, with minimum mean assortativities of approximately 0 (i.e., no attribute assortativity) at approximately 50% non-differential misclassification. For example, in practice the analysis of 50% gender non-differential misclassification changed random samples of 50% of male participants (N=117) to female and 50% of female participants to male (N=161), resulting in a new gender population distribution (50% male, 50% female). The observed mean gender assortativity of 10,000 resamples with this distribution was approximately 0 (SD: 0.03). At a low rate (5%) of hypothesized non-differential misclassification, the observed mean resampled gender assortativity was 0.14 (SD: 0.02), or 19.5% lower than the measured gender assortativity. For ethnicity, there was an even larger decrease (50.4%) in attribute assortativity with 5% non-differential misclassification of Hispanic identity. This resulted from the highly disproportionate number of self-identified non-Hispanic participants (95% of participants with non-missing ethnicity data); resampling and reclassifying 5% of both groups, in effect, doubled the network's Hispanic population, therefore having a larger impact on the measured ethnicity assortativities.

We also assessed the impact of potential, hypothesized differential misclassification by alcohol use, or underreporting of alcohol use on the baseline survey by non-drinkers, among the 423 (72%) participants who responded to the alcohol use item on the baseline survey. As shown in Figure 4.9, it appears that the observed alcohol assortativity may overestimate the effects of alcohol use on network clustering given a range of differential misclassification. Assuming that misclassification, or underreporting, was randomly distributed through the underage non-drinking population and incrementally increasing differential misclassification levels results in a parabolic trend in observed alcohol assortativities, centered at approximately 65%

misclassification by non-drinkers (under 21 years), which corresponds to 70% alcohol use in the population, with a minimum mean alcohol assortativity of 0.05 (SD: 0.03), or approximately 75% below the observed alcohol assortativity given the original distribution of alcohol use. In in the range of expected drinking levels in the population (approximately 60-70% (213), corresponding to approximately 55-65% misclassification by non-drinkers (under 21)), the generated alcohol assortativity values ranged from 0.07 (SD: 0.04) to 0.05 (SD: 0.03). When no assumptions about associations between age and misclassification were made, i.e., all non-drinkers were equally likely to under-report their alcohol use, the generated alcohol assortativities in this randomization scheme steadily decreased with increasing proportions of drinkers in the population, with a minimum mean value of 0 (SD: 0.04). In the range of expected drinking levels in the population (approximately 60-70% (213)), the generated alcohol assortativity values ranged from 0.08 (SD: 0.04) to 0.06 (SD: 0.04). In combination, these results suggest that the inclusion of the 23 non-drinkers who were over 21 years of age in the randomization scheme had a measurable impact on alcohol assortativity. In other words, this small segment of the non-drinking population may somehow drive the overall network alcohol use assortativity.

### 4.3.4 Attribute assortativity uncertainty

*Jackknife models*

Figures 4.10-4.12 show the calculated 95% confidence intervals (CI) for each of the three observed attribute assortativities for the range of tested randomly removed (jackknifed) edges or nodes (1 individual node or edge, 5-95% at 5% intervals). As expected, for all three covariates, both the jackknife removal of single nodes or edges resulted in negligible 95% confidence intervals. Note that for each variable, there appears to be a threshold for the percentage of jackknifed nodes after which no confidence intervals were assessed; this occurred when the random removal of nodes resulted in a network with an undefined attribute assortativity for the variable of interest. This result occurred at lower percentages for variables where one group was significantly smaller than the other (e.g., ethnicity, with 5% Hispanic and 85% non-Hispanic nodes), as removing a large percentage of nodes tends to remove all or most of the smaller group.

Generally, the jackknifed node removal procedures produced wider confidence intervals than jackknifed edge removals. This outcome is a logical result, given topological effects of node versus edge deletion in graphs. The width of alcohol use, ethnicity, and gender assortativity confidence intervals increased with increasing removal percentages, with the exception of 70% edge or node removal, which had slightly narrower confidence intervals than those seen with 65% removal. However, the subsequent 75% removal resulted in confidence intervals that were larger than all tested lower removal percentages.

*Bootstrapped models*

The bootstrapped edge removal scheme resulted in confidence intervals roughly equivalent to those with the 50% jackknifed edge removal. This suggests that random edge resampling with replacement on the digraph was in effect removing approximately 50% of the network edges during each resampling.

### 4.3.5 Attribute assortativity hypothesis testing

*Node null models*

As shown in Figure 4.13, the distributions of attribute assortativities across the node null models for each covariate (alcohol use, ethnicity, and gender), resulted in the rejection of the null model or hypothesis, as less than 5% of the distributions were as or more extreme than the observed assortativities. Therefore, we concluded the alcohol use, ethnicity, and gender assortativities were not generated by random chance, and represent significant preferential attachment based on each of the three individual covariates. In terms of the generated distribution of attribute assortativities for each covariate, the alcohol use and gender assortativities were fairly symmetric around 0 (no assortativity) and the ethnicity assortativity was slightly skewed towards positive assortivities. Overall, the modeled values spanned a range of values approximately twice the size of the originally observed attribute assortativities. This means that the model produced meaningful positive and negative values in non-negligible frequencies. Although 95% of the analyzed null model attribute assortativities were at or below the observed attribute assortativity, these results highlight the fact that it is indeed possible for networks to exhibit attribute assortativity by mere chance.

*Edge null models*

Similarly, as shown in Figure 4.14, the distributions of attribute assortativities across the edge null models resulted in the rejection of the null model or hypothesis, as less than 5% of the attribute assortativity distributions were as or more extreme than the observed attribute assortativities. Therefore, we concluded nodes were preferentially connected based on the wholistic combination of individual participant covariates (both measured and unmeasured). In other words, the alcohol use, ethnicity, and gender assortativities were not generated by random chance, and represent significant preferential contact based on similarity across multiple individual characteristics. In terms of the generated distribution of attribute assortativities, the alcohol use and gender assortativities were centered and symmetric around 0 (no assortativity), whereas the ethnicity assortativity distribution was slightly skewed towards positive values. Overall, the modeled attribute assortativity distributions spanned ranges approximately twice the size of the originally observed attribute assortativities.

## 4.4 Discussion

We assessed three attribute assortativities of the measured eX-FLU baseline social network (alcohol use, ethnicity, and gender), and developed and implemented multiple approaches for hypothesis testing and analyzing the potential impacts of imperfect individual and network data. Specifically, the sensitivity analyses and uncertainty quantification systematically explored the robustness of attribute assortativity in the face of multiple types of measurement error. For the first time (that we are aware of), we addressed how measurement error and nonresponse of self-reported covariates may bias attribute assortativity. We then used similar approaches to quantify the range of values in which the network's "true" attribute assortativity likely lies, given a range of levels of random network node and edge data error. Overall, we found that attribute assortativity biases and uncertainty may be predictable, given an *a priori* defined amount and type of data error, although how these errors would be effectively quantified requires further investigation. Nonetheless, these methods provide a framework for attribute assortativity uncertainty analyses, as well as hypothesis testing, on sociocentric networks, particularly in contact network studies.

We outlined and implemented sensitivity analyses to better understand the potential impacts that missing or inaccurate measurements of participant covariates could have on the observed network properties, specifically the network's attribute assortativity, or preferential attachment between similar nodes. We found that generally, nonresponse did not seem to have a notable effect on the observed attribute assortativity whereas both differential and non-differential misclassification of the covariate of interest appears to bias attribute assortativity results away from the null (i.e., an attribute assortativity of 0) in smooth and predictable ways. Additionally, we found that interactions between centrality and individual-level data quality may impact assortativity results in meaningful ways.

In terms of nonresponse effects, we found that if nodes with missing data are randomly assigned a characteristic, we did not see meaningful change in assortativities. However, we note that in our network, missingness did not occur randomly across nodes; there were clear associations between nonresponse and centrality measures, especially outdegree. This suggests that using a multiple imputation approach (based on individual covariate and/or centrality (i.e., one or multiple centrality metrics) instead of random assignment may be appropriate for networks with differential nonresponse in covariate and relational data. Given that outdegree is a unique centrality measure in that it is self-reported, as opposed to indegree, for example, which is dependent on other participants' reporting, outdegree can be seen as a proxy for participation or survey engagement. In this case, nodes with missing personal data were also likely to have reported fewer, if any, contacts on the contact survey. Therefore, nodes with incomplete or missing personal data also tended to contribute less to network structure than nodes who responded to both the contact and baseline surveys. Therefore, their inclusion or exclusion in assortativity calculations would appear to have less of an effect on the network assortativity than people with complete data (who typically have higher outdegree). This is particularly true of isolated nodes, who had no connections to other participants, as isolates are not included in assortativity calculations. Therefore, even with imputed characteristics, these participants will not factor into any assortativity calculations in this particular sensitivity analysis procedure. Therefore, analyzing and acknowledging any differential centrality by nonresponse status is important when conveying these types of results.

In addition to nonresponse effects, we developed sensitivity analysis procedures to understand how participant covariate measurement error might affect inference. We first assessed the observed positive assortativities' robustness in light of potential non-differential misclassification by ethnicity or gender. We analyzed these possible non-differential misclassification scenarios not due to any hypothesized actual measurement error, but merely as an illustration of the outlined methods. We then analyzed potential differential mis-classification of non-drinkers, which could explain this study population's low levels of reported alcohol use. In the case of any misclassification, regardless of type, the sensitivity analyses showed that the observed assortativities may have been biased away from the null and may have overestimated the true assortativity. It appears that the amount of bias is a result of hypothesized type and level of misclassification, in addition to the measured distribution of a given characteristic and any centrality differences between groups.

The results of the non-differential misclassification sensitivity analyses highlight the effect of observed distributions on the potential impacts of misclassification. The network was similarly assortative by ethnicity and gender, but given just 5% non-differential misclassification, ethnicity assortativity was reduced by half whereas gender was reduced by only 20%. Similarly, when non-differential misclassification was in the range of 5-20%, ethnicity assortativity was reduced by between 50-90%, whereas gender was only reduced by 20-63%. The different amounts of bias appear to be due to differing measured characteristic distributions; gender was somewhat evenly split between men and women whereas participants overwhelmingly self-identified as non-Hispanic. Therefore, the relative impact of low levels of non-differential misclassification on the distributions substantially impacted the bias. In effect, dramatically changing the characteristic make-up of the network, as was the case in the ethnicity analyses, changes the crude probability that dissimilar individuals will be connected by an edge. In the real world, we could say that changing the ethnic make-up of the network by increasing Hispanic participation generally gave non-Hispanic participants more opportunities to connect with Hispanic participants.

In the case of possible differential misclassification by self-reported alcohol consumption, we assessed how different levels of alcohol use underreporting may have impacted the observed assortativity. Although we believe that baseline alcohol use underreporting was minimal in this study population, given the consistent results across both study years (see Tables 2.2 and 2.3 in Chapter 2), social desirability of alcohol use in college, and accepted validity of self-reported drinking in this population (63,235), we still note that the proportion of the study population who reported being current drinkers at baseline was well below that seen in other research (213). However, we acknowledge that measurement error, i.e., misclassification, in the baseline alcohol use measurement could have arisen due to unfounded concerns about data privacy or negative consequences of reporting alcohol use. Therefore, while the results of the sensitivity analysis suggest that the observed alcohol assortativity in the social network may be biased or inflated, we do not see a compelling reason to reject the observed result in favor of a more conservative estimate. However, studies with *a priori* hypotheses regarding differential misclassification and/or measurement error should consider if the methods we've outlined here would bolster the strength and veracity of their analyses.

Much like the effect of associated centrality and non-response, centrality differences between different groups in the network will impact how much misclassification biases the observed assortativity. Logically, nodes with higher centralities, particularly degree, will have a have a larger contribution in any assortativity calculation, as their connections comprise larger percentages of the network's total edge count. Therefore, in a measured network with positive assortativity and differences in centrality by the characteristic of interest (as was the case with alcohol use in this study), random reclassification of nodes in a single direction should: 1) reduce the centrality differences across groups, and 2) reduce the impact that a given characteristic has on the observed assortativity. In other words, we found that differential misclassification of a characteristic with significant centrality associations appears to be an important moderating factor in how the misclassification biases the observed assortativity. We also hypothesize that if differential misclassification is itself associated with centrality in addition to a given characteristic, this could further complicate or confound the effect that the misclassification has on observed assortativity. This may explain the observed differences in the results of the sensitivity analyses including non-drinkers over 21 years and those excluding this small segment of the population (approximately 8.5% of non-drinkers). If self-reported non-drinkers over 21 years did not mis-report their alcohol use and were more central than underage non-drinkers, excluding this group from randomization and resampling would maintain a "core" of central non-drinkers. Assuming they were highly clustered among themselves, which is possible given the high observed assortativity by alcohol use, this cluster of inter-connected non-drinkers would influence assortativity regardless of misclassification among other non-drinkers. We note that given the lack of observed centrality differences by ethnicity or gender, we could not investigate how differential centrality and non-differential misclassification might interact and bias assortativity. Additionally, we acknowledge that the observed results and our conclusions may be unique to assortative networks and that misclassification may bias disassortativity in different ways, making investigation of disassortative networks a useful next step in future work. Collectively, the sensitivity analyses results suggest that not only does individual data quality impact the observed assortativity, but that the accuracy of degree measurements, is paramount to assortativity calculations. If nonresponse and misclassification were associated with outdegree (or more generally, degree), the error could be propagated throughout assortativity analyses.

In addition to investigating covariate measurement error and its effects on observed attribute assortativity, we explored methods to quantify attribute assortativity uncertainty resulting from underlying network data uncertainty.  We repeated uncertainty calculations across a range of resampling proportions as well as by the resampled network structure (nodes or edges). Although resampling proportions above 50-60% are unusual, at least at the size of the analyzed network, we performed the analyses across the range of 0-95% for illustrative purposes. Much like the sensitivity analyses, the uncertainty calculations are based on the measured network and thus rely on covariate and network data completeness and accuracy. As jackknife resampling samples the measured network in an attempt to mirror the original measurement's sampling of the "true" underlying network, its results are based in the assumption that the measured network was itself randomly sampled. However, no study is truly the result of random population sampling, and network studies, including the eX-FLU study, often use snowball sampling for enrollment. Therefore, future developments and refinements of the uncertainty analysis procedure could incorporate this fact. Indeed, one could resample the network by sampling nodes in an iterative process, whereby the first sampled node is random and subsequent sampled nodes are directly connected to the previously sampled node. This would then simulate the snowball process and account for the original sampling procedure.

Overall, we found that node removal or jackknifing provided more conservative confidence interval estimates than edge removal or jackknifing. This is a logical result, as node removal also removes edges from the network, so it in effect combines the two error mechanisms. In contrast, edge removal serves to reduce random nodes' degrees and connectivity in the network, but not the overall network size (number of nodes). We can speculate then that in a sparse (low density) network, low levels of edge or node removal would have a more marked effect on assortativity than in a dense network. In fact, the width of the confidence interval of node and edge jackknifed resampling depends on resampling proportions, network size, density, as well as degree distribution. As size and density increase, the impact of a given amount of node or edge removal on uncertainty would lessen. Similarly, as previously noted by Cohen et al. (287), scale-free (e.g., right skewed degree distribution) networks are resilient to node removal, up to the point where the probability of removing a highly connected node becomes high enough, therefore disrupting the global network structure.  However, if most nodes have a degree similar to the network's mean degree (e.g., a random network), node removal at lower levels would result in wider confidence intervals.

We also examined bootstrapped edge resampling, or resampling with replacement from the observed distribution of edges; however, on simple graphs resampling edges with replacement will typically result in a multigraph, which is not comparable to the simple graph, especially with regards to assortativity. In this case, any dyad could contain any number of edges (up to the number in the original network). Thus in each resampled network, there exists the possibility that a small sample of connected nodes could disproportionately affect the assortativity, as they would hold a disproportionate proportion of the network's edges. In multigraphs with unconstrained maximum edges per dyad, the bootstrapping procedure is distinct from jackknifing and may be the most reasonable method for calculating uncertainty. In fact, generally, if one cannot make an *a priori* hypothesis about level of accuracy in an observed network (which we, as others have, note that this is usually impossible (172)), the "bootstrapped" (i.e., random percentage edge removal) resampling method may be most appropriate because it does not use pre-defined error levels.

We identified a number of studies that examined the robustness of a variety of network characteristics (e.g., degree distribution), given hypothetical network measurement and sampling error (57,181,205,206,209,273–275,286,288,289). We are, however, aware of only one previous study that systematically looked at the sensitivity of assortativity given potential network measurement error. Kossinets investigated the impact of missing data on a number of characteristics, including degree assortativity, or the tendency of a node's degree to be similar to that of their neighbors (209). This study found two different network data error types (isolated nodes due to non-response and underreported numbers of contacts) bias degree assortativity towards the null, but that overall, different types of network data error can bias degree assortativity in opposite directions. Similarly, Newman has suggested that degree assortative networks are generally robust to node removal (57). However, it may be that attribute assortativity, as opposed to degree assortativity, has different reactions to measurement error, given the aforementioned complicated relationships between degree and isolation, individual attributes, and individual attribute data accuracy. Despite this body of literature on the robustness of network statistics, we are unaware of any studies that quantify uncertainty and confidence limits using resampling methods. Indeed, Franz et al. (2009) noted that generally, social network analysis is currently limited by an inability to determine valid confidence limits for network statistics (172). Our exploration into quantifying the reliability of attribute assortativity is a useful extension of these previous robustness studies that provides a framework for addressing this limitation.

We also described and implemented two null models to test the significance of assortativity, or the probability that the measured network's observed assortativity was not the result of chance connections between similar individuals, but instead a result of selection, whereby edges were preferential by a given characteristic. In doing so, we highlight the need to clearly define and articulate the null model and its implementation in order to draw appropriate conclusions from results. Indeed, as noted by Dormann, communicating the null model's structure and implementation is crucial for critical reviews of the model's assumptions and results, as well as reproducibility (280). Appropriate null models can account for any biases in the observed data (179,180), but incorrectly defined or implemented models may lead to biased and spurious conclusions (286). Surprisingly, while the field of ecology has generally adopted using null models in animal network analyses, it has not been widely embraced in other fields of social network research (152). This may be due to disinterest in generalizable results, as some social network analysis frameworks lean towards describing individual networks in isolation, as opposed to making inferences about populations of networks or the underlying “true network” (277). Thus, if a study’s aim to merely describe the observed network, as opposed to attempting to infer larger and generalizable patterns from the sampled network, comparisons to null models (as well as sensitivity and uncertainty analyses), are moot. If greater understanding of the underlying social structures and mechanisms that gave rise to the sampled network are the research goal, then incorporating null models into analytic plans is imperative. Indeed, we must question all conclusions drawn using poorly defined models, especially those without detailed descriptions of the model aim and implementation.

Given that random networks can exhibit assortative mixing (155) and sampling and data collection methods can lead to overestimated assortativities (152), it is important to attempt to isolate the hypothesized selection effects from random effects. Based on the definition and implementation of our null models, we assert that if the goal of an attribute assortativity hypothesis test is to isolate the effects of a single characteristic on network edge formation, the node null model is the more appropriate model. As much as possible given the observed data, this model reflects the sole effect of this characteristic on network connections. In contrast, the edge null model tests the hypothesis that the combination of all node characteristics, both known and unknown, are responsible for an observed attribute assortativities.

The node model randomizes a given characteristic across nodes while maintaining network topology. In doing so, the model inherently assumes that the network structure may be dependent on all other node characteristics and thereby isolates the effect of a single trait on selection (by only randomizing that variable). Indeed, as edges and nodes remain constant across resamples, network homophily by other covariates (i.e., other than the randomized variable) is controlled for in this null model. Although this null model appears to be the most effective way to isolate the effects of a single variable on selection, the node null model is not without considerations. Selection likely occurs across multiple traits and their interactions (59); the node null model breaks the association between a covariate of interest and any and all associated traits (e.g., an association between age and alcohol use). Therefore, this null model should control for confounding by latent homophily, but the results and their interpretation exclude any possible covariate interactions that drive selection. Notably, this null model and inferences rely on covariate data quality and reliability. As shown in the sensitivity analyses, attribute assortativity can be quite sensitive to individual covariate measurement error. Therefore, when resampling maintains a covariate's observed distribution overall, any covariate measurement error will propagate throughout the null models and could bias the conclusions. By including nodes with missing data for a given covariate in resampling, we believe that the node null model controls for nonresponse and limits this particular potential effect. However, the results of the covariate attribute assortativity sensitivity analyses show that misclassification can materially bias attribute assortativity and thus we suggest further investigations into how to incorporate covariate data uncertainty in null models.

Unlike the node null model, the edge null model is not based on the observed or measured network edges (although it does rely on the sampled, or observed, nodes, i.e., network boundary). Instead, the edge null model randomly samples edges from a set of all potential edges. In practice, this isolates some but not all other network properties. While the density of the network will remain constant across all resamples, the degree distribution will change and therefore and any associations between individual characteristics and degree will be broken (as will a number of other network properties). We can then conclude that this type of edge model might be used for hypothesis testing on networks with known associations between multiple node characteristics, or if there is an hypothesized interaction relevant to selection. However, we acknowledge that, as with the node null model, covariate measurement inaccuracy may bias observed attribute assortativities on the null models, and therefore any conclusions.

Much like propagated biases from covariate measurement error, these hypothesis tests can also be affected by systematic issues in network data quality, reliability, and sampling. In particular, inferences are intrinsically linked to the network's sampling and measurement. In particular, network boundaries drawn (e.g., a set of participants enrolled in a study) with heavily skewed covariate distribution(s) (that differ from the underlying population distribution(s)) may result in biased resampled null models for attribute assortativity (akin to selection bias in epidemiological studies). In terms of sampled, or measured, network edges, we can imagine scenarios in which reported contacts might be prone to reporting and/or recall bias, if individual traits influence the propensity to remember or report certain relationships or contacts. For example, elderly individuals may have more difficulty with recalling contacts than younger participants. Similarly, shared (or perhaps disparate) traits with other network members may effect individuals' contact reporting behavior. For instance, younger participants may more accurately recall their contacts with other young participants than elderly network members, or individuals may be less likely to report sensitive relationships (e.g., a sexual relationship with a co-worker). Any one, or a combination of, these biases present in the data collection will result in null models that are biased away from a truly random network. Generally, these potential biases, as well as any caveats and assumptions related to data collection and sampling, should be noted when defining and making any inferences based on these (or any) attribute assortativity null models (as they should be in all studies).

Given that the implemented analyses and results are presented here as illustrations of the outlined approaches, we will for the most part restrict our noted limitations to the approaches and not delve into the limitations of the dataset, measured network, and the observed attribute assortativity results. However, a clear limitation of the analyzed dataset and the selected covariates is that we were unable to investigate how observed disassortativity might be impacted by data quality and completeness. It is quite possible that disassortative networks have different properties that would change the conclusions we have drawn. Additionally, while we have addressed possible errors individually, we must note that the observed network and covariates (as with all observed data) likely contains a mix of unknown covariate and network measurement errors (172), including systematic sampling errors (206). As discussed, any biased data or sampling will propagate throughout analyses and affect inference and generalizability, and we suggest that future work could delve into how to appropriately combine the outlined approaches to account for multiple error sources.

For the first time that we are aware of, we have outlined and defined multiple attribute assortativity sensitivity analyses, uncertainty quantifications, and null model-hypothesis testing approaches and applied these methods to a measured social network of undergraduate students. In toto, the results of these analyses show the potential impacts that data quality, measurement error, and the measured network can have on observed assortativity. We suggest that it should be standard practice to present sensitivity analysis for assortativity (and other network properties), and to hypothesize possible confounding or bias related to data quality and completeness. We also strongly suggest that researchers make a point of clearly defining and describing the methods they have employed in their social network analyses. Unclear or undefined resampling methods and inappropriately defined or implemented null models can lead to incorrect conclusions and inferences, ultimately making generalizability and duplication difficult or impossible.

## 4.5 Tables

**Table 4.1** Observed properties of the eX-FLU baseline simple digraph social network measured via contact survey (2013).

| **Structure or property** | |
|---|---:|
| Nodes[a] | 590 |
| Directed edges[a] | 1275 |
| Dyads[a] | 988 |
| Dyad reciprocity (global)[b] | 0.29 |
| Density | 0.023 |
| Transitivity[c] | 0.68 |
| Isolates[a] | 122 (0.21) |
| Degree[d,e] | 3.35 (3.97) |
| Indegree[d,f] | 2.16 (2.57) |
| Outdegree[d,g] | 2.16 (3.38) |
| Dyad reciprocity (node)[d,h] | 0.27 (0.32) |

[a] N or N (%)
[b] Proportion of all network dyads containing reciprocal edges.
[c] Not defined for digraphs, computed using simple (dyad) graph.
[d] Mean (standard deviation (SD)).
[e] Unique neighbors connected to a participant by at least one edge.
[f] Unique neighbors who named a participant as a contact.
[g] Unique neighbors a participant named as a contact.
[h] Proportion of neighbors connected to a node by reciprocal edges.

**Table 4.2** Social network characteristics of self-reported drinkers, non-drinkers, and those with missing alcohol use data in the eX-FLU study (2013), as observed in the baseline social network.

| | **Drinker** | **Non-drinker** | ***P value***[b] | **Missing alcohol use data** | ***P value***[c] |
|---|---|---|---|---|---|
| N | 155 (0.26) | 268 (0.45) | | 167 (0.28) | |
| Isolates | 37 (0.14) | 18 ( 0.12 ) | | 67 (0.40) | |
| Degree[d] | 5.18 (5.14) | 3.5 (3.56) | <0.001* | 1.40 (1.98) | <0.0001* |
| Indegree[e] | 3.34 (3.49) | 2.08 (2.13) | <0.0001* | 1.19 (1.59) | <0.0001* |
| Outdegree[f] | 3.61 (4.44) | 2.43 (3.01) | <0.01* | 0.38 (1.50) | <0.0001* |
| Dyad reciprocity (node)[g] | 0.32 (0.31) | 0.29 (0.34) | 0.81 | 0.06 (0.19) | <0.0001* |

*P<0.05

[a] Data are N (%) or Mean (Standard deviation (SD)).

[b] Difference between drinker and non-drinker distributions; P-values calculated by resampled difference in means tests.

[c] Difference between distributions for participants with and without alcohol use data; P-values calculated by resampled difference in means tests.

[d] Unique neighbors connected to a participant by at least one edge.

[e] Unique neighbors who named a participant as a contact.

[f] Unique neighbors a participant named as a contact.

[g] Proportion of neighbors connected to a node by reciprocal edges.

**Table 4.3** Social network characteristics participants who self-identified as Hispanic, non-Hispanic, and those without ethnicity data in the eX-FLU study (2013), as observed in the baseline social network.

| | **Hispanic** | **Non-Hispanic** | ***P value*[b]** | **Missing ethnicity data** | ***P value*[c]** |
|---|---|---|---|---|---|
| N | 27 (0.05) | 505 (0.86) | | 58 (0.10) | |
| Isolates | 3 (0.11) | 100 (0.20) | | 19 (0.33) | |
| Degree[d] | 4.30 (3.68) | 3.44 (4.06) | 0.30 | 2.09 (2.87) | 0.01* |
| Indegree[e] | 2.67 (2.78) | 2.20 (2.60) | 0.38 | 1.55 (2.04) | 0.06 |
| Outdegree[f] | 2.74 (3.31) | 2.28 (3.44) | 0.51 | 0.90 (2.48) | <0.01* |
| Dyad reciprocity (node)[g] | 0.28 (0.32) | 0.28 (0.33) | 0.97 | 0.10 (0.21) | <0.001* |

*P<0.05

[a] Data are N (%) or Mean (SD).

[b] Difference between Hispanic and non-Hispanic distributions; P-values calculated by resampled difference in means tests.

[c] Difference between distributions for participants with and without ethnicity data; P-values calculated by resampled difference in means tests.

[d] Unique neighbors connected to a participant by at least one edge.

[e] Unique neighbors who named a participant as a contact.

[f] Unique neighbors a participant named as a contact.

[g] Proportion of neighbors connected to a node by reciprocal edges.

**Table 4.4** Social network characteristics participants who self-identified as male, female, and those without self-reported gender data in the eX-FLU study (2013), as observed in the baseline social network.

| | **Male** | **Female** | ***P value***[b] | **Missing gender data** | ***P value***[c] |
|---|---|---|---|---|---|
| N | 235 (0.40) | 323 (0.55) | | 32 (0.05) | |
| Isolates | 47 (0.20) | 61 (0.19) | | 14 (0.44) | |
| Degree[d] | 3.41 (3.97) | 3.5 (4.07) | 0.79 | 1.31 (1.98) | <0.001* |
| Indegree[e] | 2.15 (2.52) | 2.26 (2.64) | 0.64 | 1.28 (1.97) | 0.03* |
| Outdegree[f] | 2.14 (3.42) | 2.38 (3.44) | 0.42 | 0.03 (0.17) | <0.0001* |
| Dyad reciprocity (node)[g] | 0.21 (0.27) | 0.33 (0.35) | <0.0001* | 0 (0) | <0.0001* |

*P<0.05

[a] Data are N (%) or Mean (SD).

[b] Difference between male and female distributions; P-values calculated by resampled difference in means tests.

[c] Difference between distributions for participants with and without gender data; P-values calculated by resampled difference in means tests.

[d] Unique neighbors connected to a participant by at least one edge.

[e] Unique neighbors who named a participant as a contact.

[f] Unique neighbors a participant named as a contact.

[g] Proportion of neighbors connected to a node by reciprocal edges.

## 4.6 Figures

**Figure 4.1** Simple digraph visualization of the baseline (week 1) social network of reported face-to-face contacts of undergraduate students in the eX-FLU study (N=590; 2013). Nodes (i.e., participants) are represented as circles and each directed edge (i.e., line) represents a reported contact between participants on the contact survey, with an arrow pointed towards the reported contact. Reciprocal contacts reported by both participants are represented by double arrows.

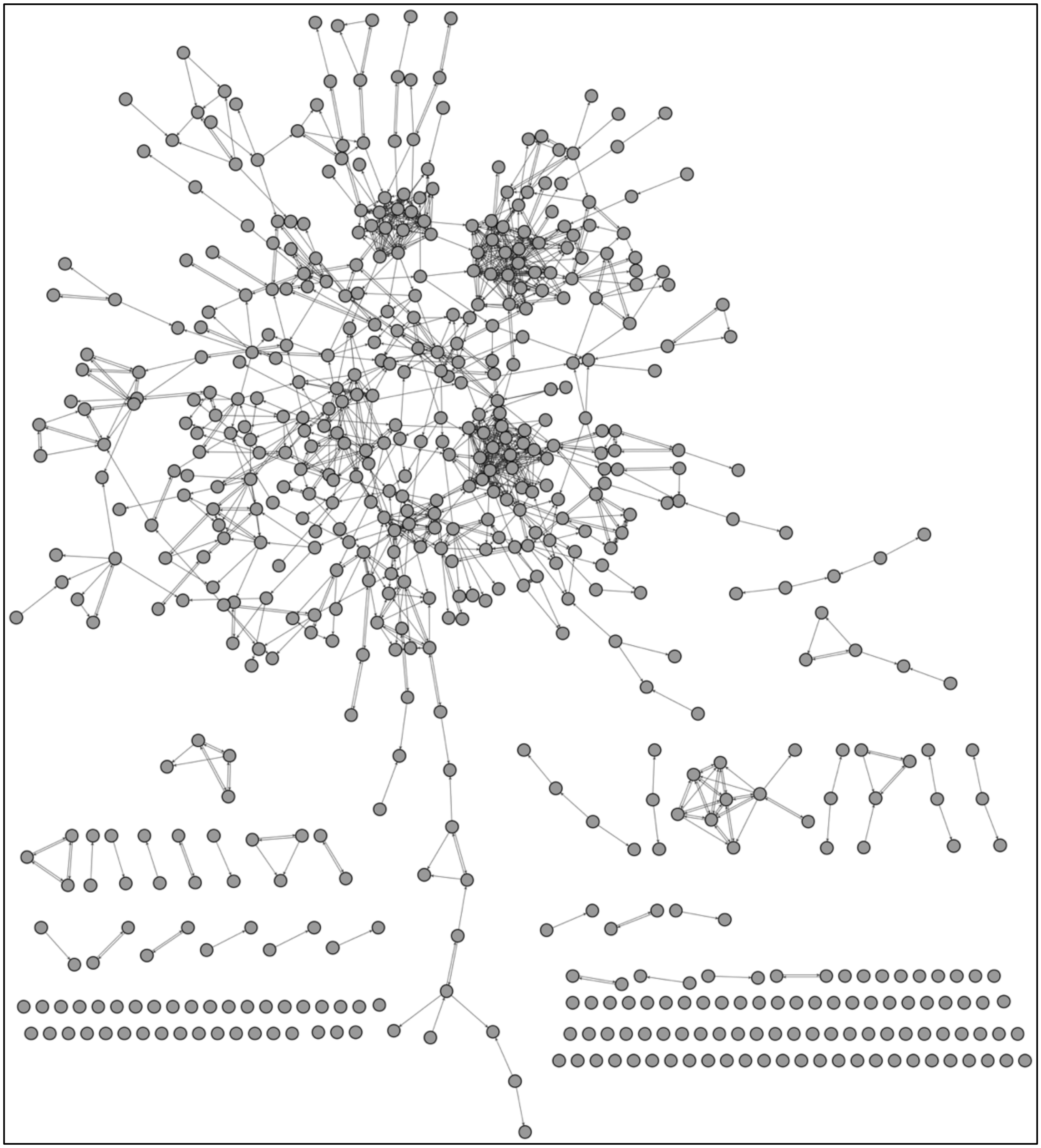


**Figure 4.2** Alcohol assortativity in the simple digraph visualization of the baseline (week 1) social network of reported face-to-face contacts of undergraduate students (2013). Panel A shows all enrolled participants colored by alcohol use and nonresponse status. Panel B contains only the 423 (72%) participants with alcohol use data. Nodes (i.e., participants) are represented as circles, colored according to alcohol use; participants who declined to report their alcohol use are represented by the smaller gray circles. Each directed edge (i.e., line) represents a reported contact between two participants on the contact survey, with an arrow pointed towards the reported contact. Reciprocal contacts reported by both participants are represented by double arrows.

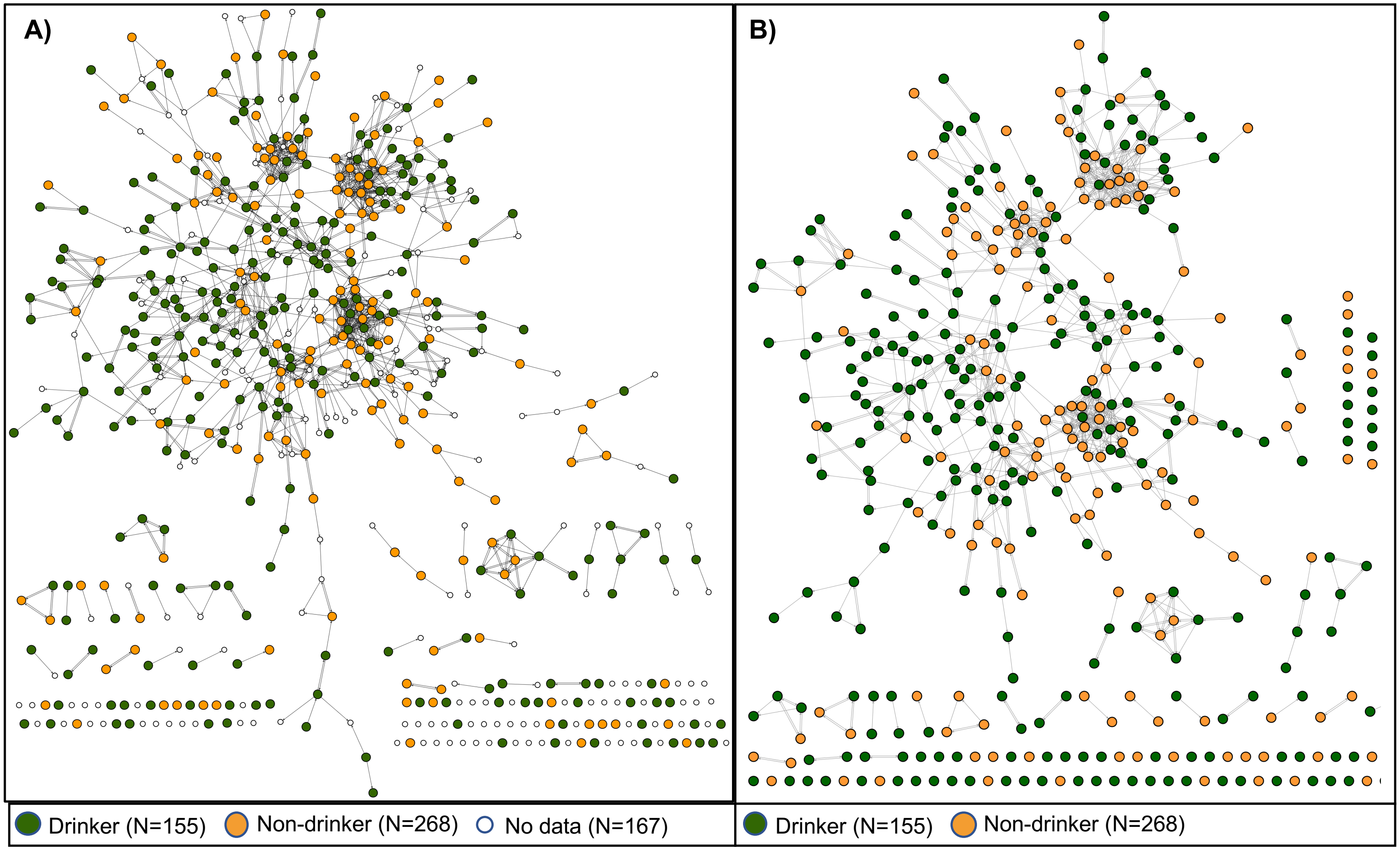


**Figure 4.3** Ethnicity assortativity in the simple digraph visualization of the baseline (week 1) social network of reported face-to-face contacts of undergraduate students (2013). Panel A shows all enrolled participants colored by ethnicity and nonresponse status. Panel B contains only the 532 (90%) participants with ethnicity data. Nodes (i.e., participants) are represented as circles, colored according to ethnicity; participants who declined to report their ethnicity are represented by the smaller gray circles. Each directed edge (i.e., line) represents a reported contact between two participants on the contact survey, with an arrow pointed towards the reported contact. Reciprocal contacts reported by both participants are represented by double arrows.

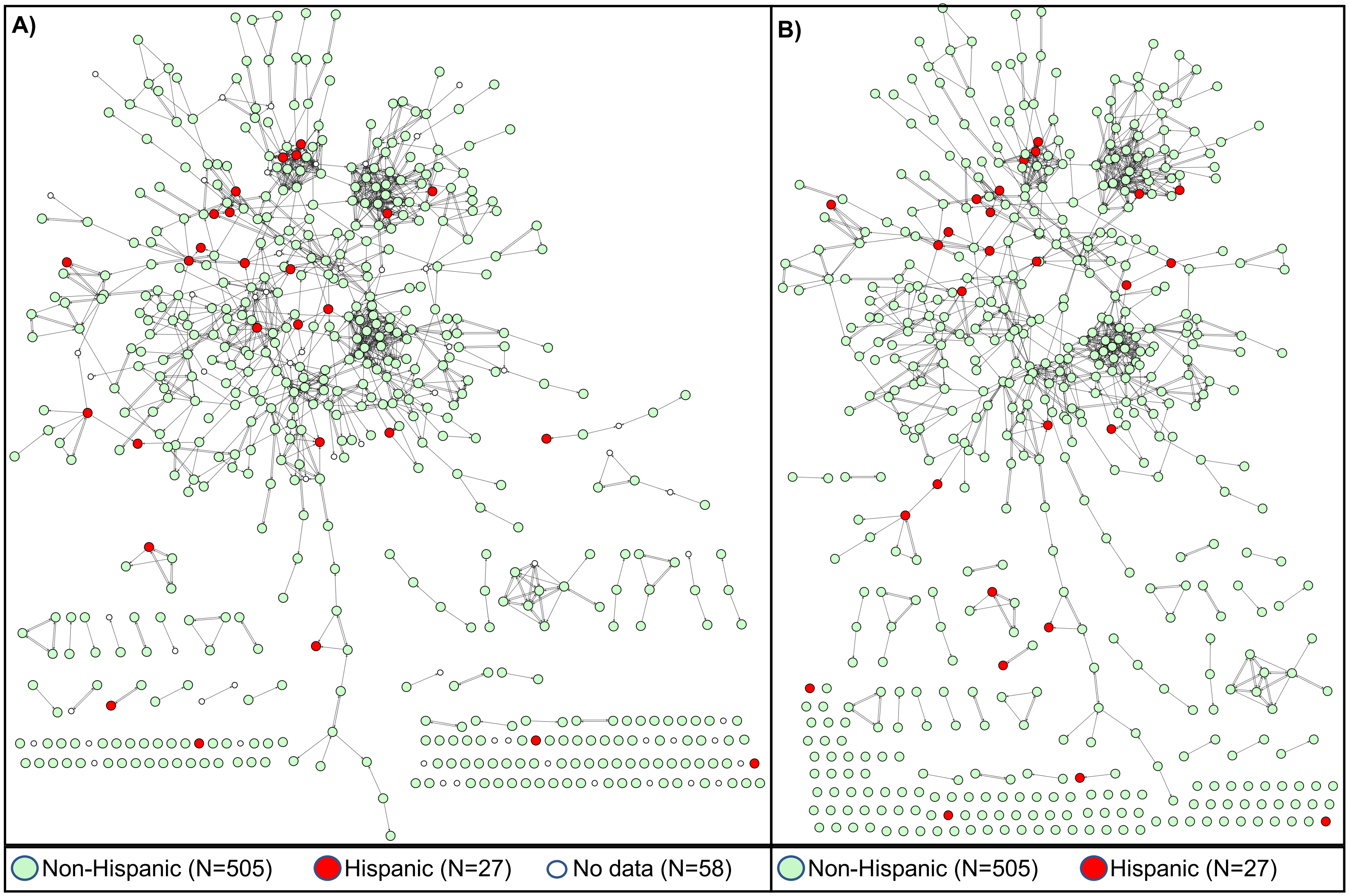


**Figure 4.4** Gender assortativity in the simple digraph visualization of the baseline (week 1) social network of reported face-to-face contacts of undergraduate students (2013). Panel A shows all enrolled participants colored by gender and nonresponse status. Panel B contains only the 558 (95%) participants with ethnicity data. Nodes (i.e., participants) are represented as circles, colored according to gender; participants who declined to report their gender are represented by the smaller gray circles. Each directed edge (i.e., line) represents a reported contact between two participants on the contact survey, with an arrow pointed towards the reported contact. Reciprocal contacts reported by both participants are represented by double arrows.

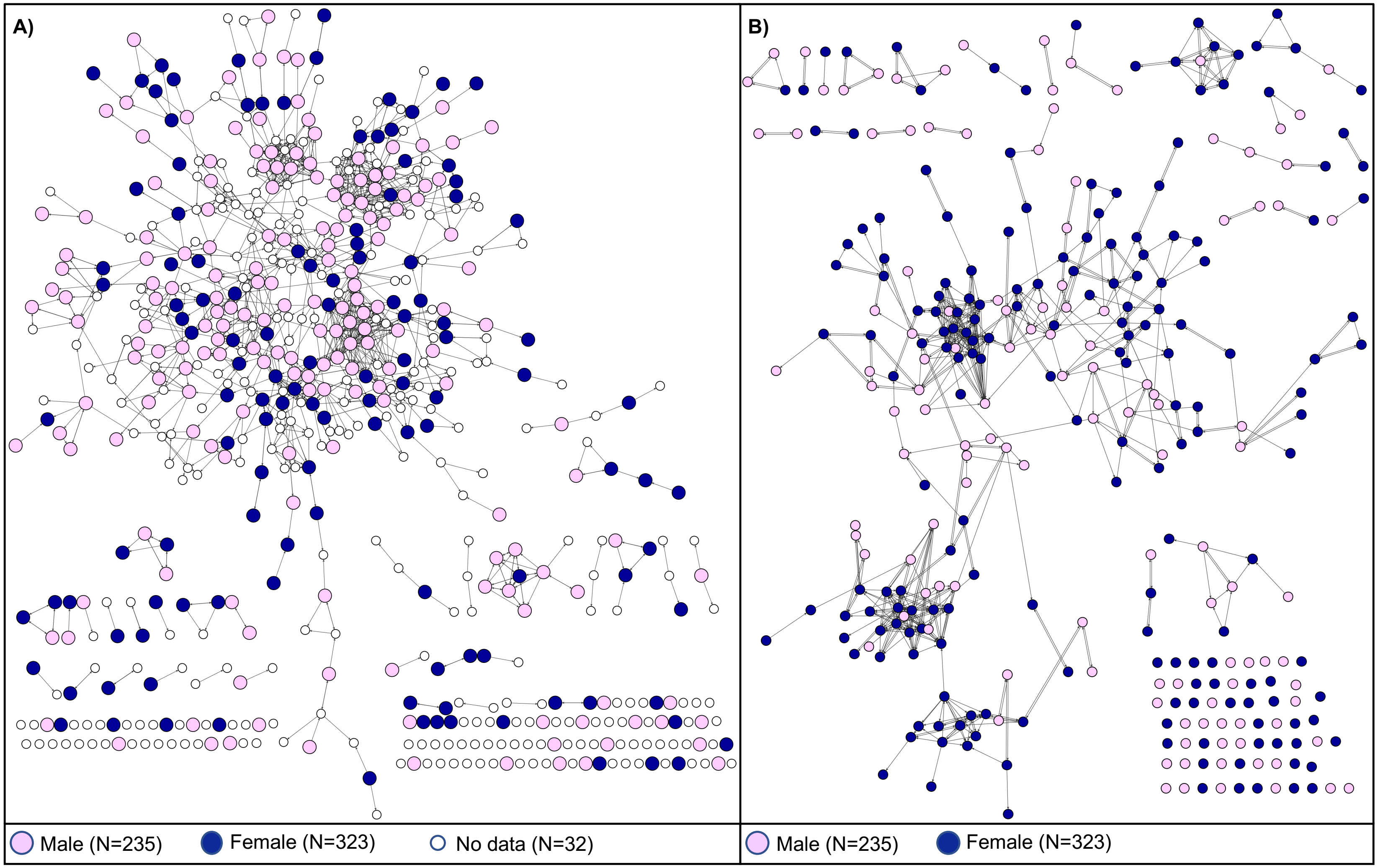


**Figure 4.5** Results of the alcohol use assortativity nonresponse sensitivity analysis (2013). The blue diamond corresponds to the observed attribute assortativity with 28% nonresponse and the observed proportion of drinkers in the study (26%; non-drinkers: 45%). Each red point and whiskers corresponds to the mean (SD) alcohol assortativity on the network after randomizing alcohol use data onto nodes with missing data on the baseline survey (10,000 randomizations). The thin blue dotted line at approximately 35% drinkers in the study population corresponds to the observed prevalence of drinking among participants with alcohol use data. The thick dotted line on the right side of the figure represents the expected prevalence of alcohol use based on previous undergraduate alcohol use research (or as close to the prevalence as possible given the observed data and nonresponse level).

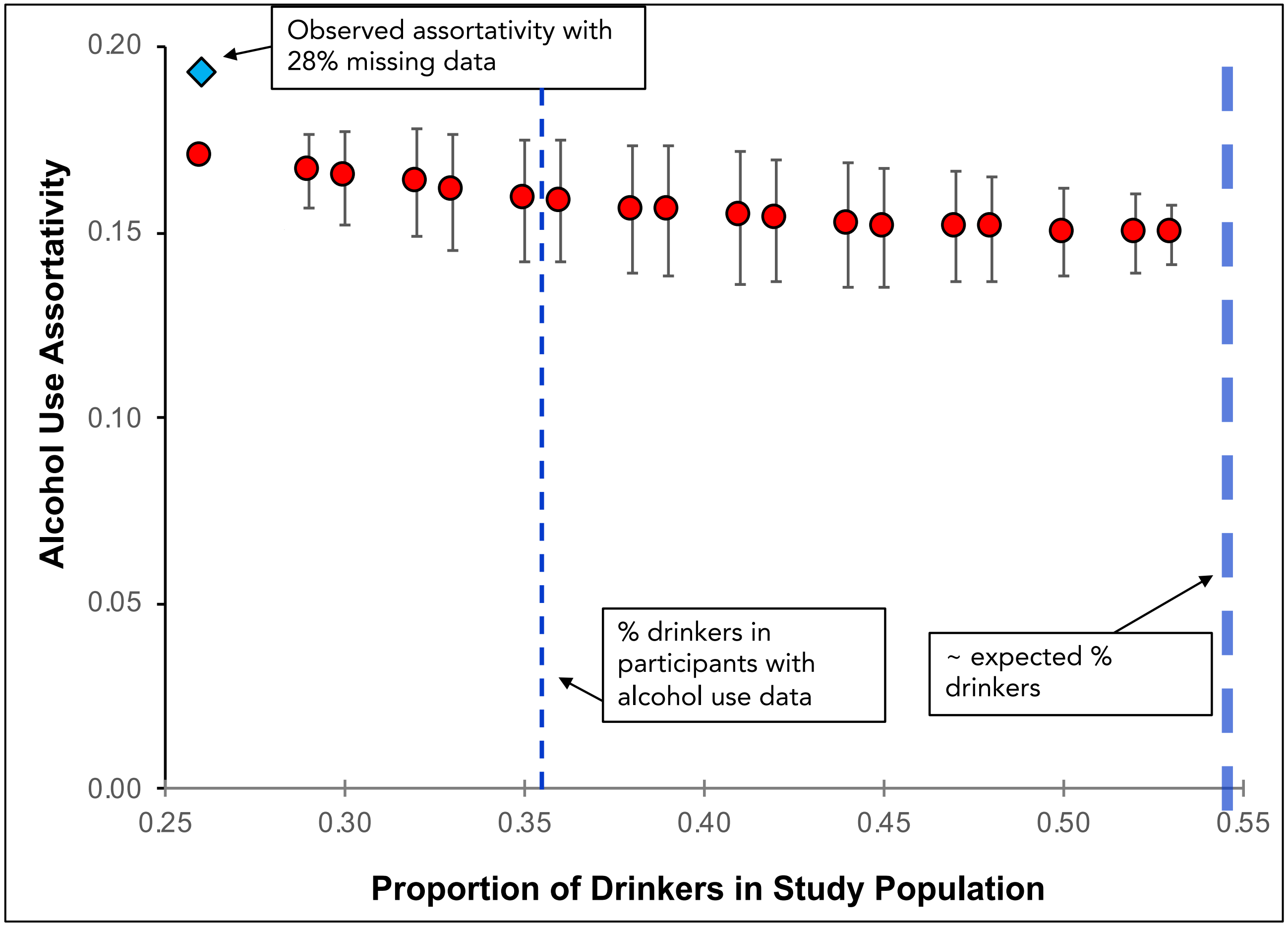


**Figure 4.6** Results of the ethnicity assortativity nonresponse sensitivity analysis (2013). The blue diamond corresponds to the observed attribute assortativity with 9.8% nonresponse and the observed proportion of Non-Hispanic participants in the study (86%; Hispanic: 5%). Each red point and whiskers corresponds to the mean (SD) ethnicity assortativity on the network after randomizing ethnicity data onto nodes with missing data on the baseline survey (10,000 randomizations).

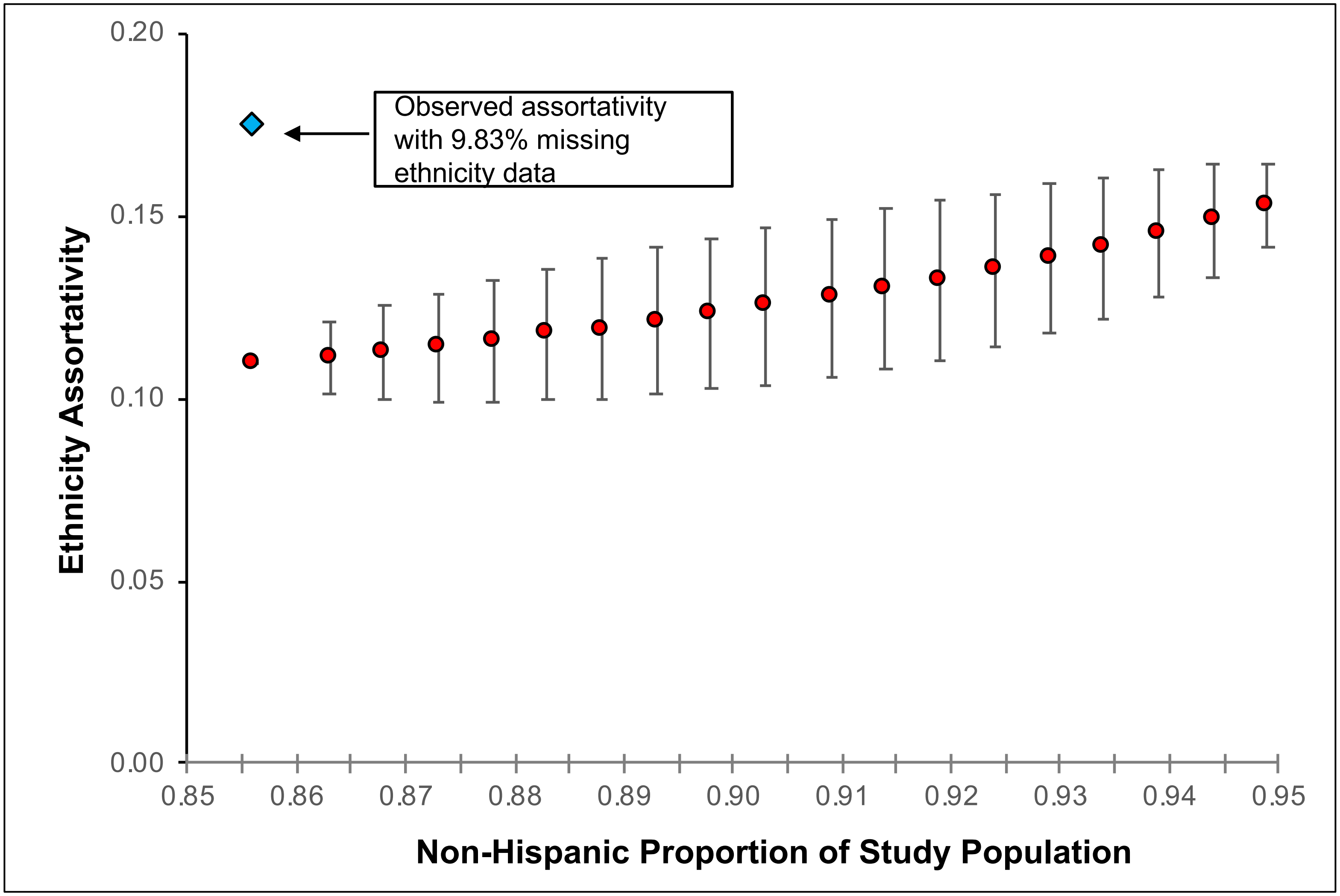


**Figure 4.7** Results of the gender assortativity nonresponse sensitivity analysis (2013). The blue diamond corresponds to the observed attribute assortativity with 5% nonresponse and the observed proportion of female participants in the study (55%; male: 40%). Each red point and whiskers corresponds to the mean (SD) gender assortativity on the network after randomizing gender data onto nodes with missing data on the baseline survey (10,000 randomizations).

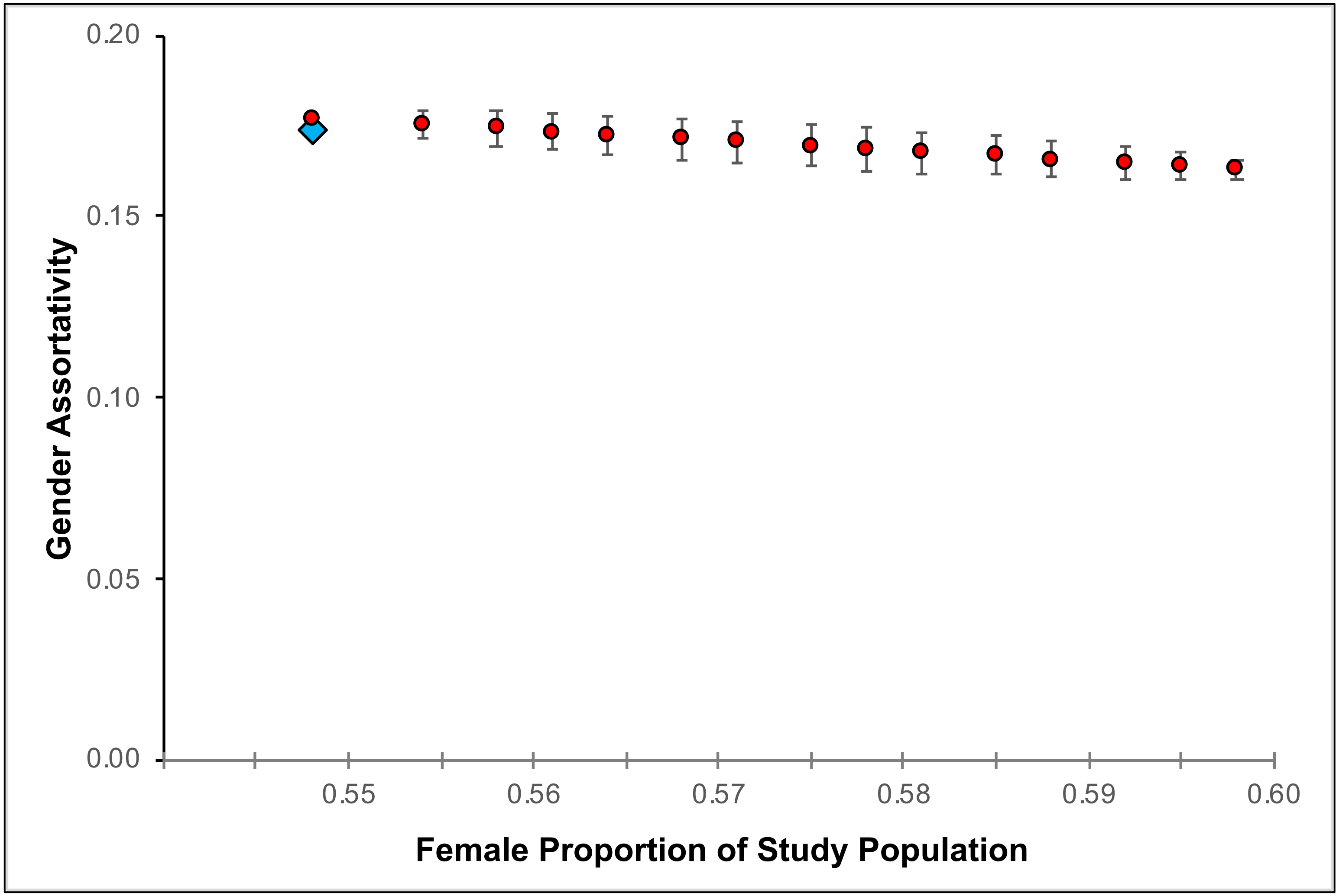


**Figure 4.8** Sensitivity analysis to assess impacts of non-differential misclassification of ethnicity (Panel A) and gender (Panel B) in the social network (2013). Each red dot and whiskers represent the mean (SD) attribute assortativity for 10,000 random samples of a given percent of non-differential misclassification.

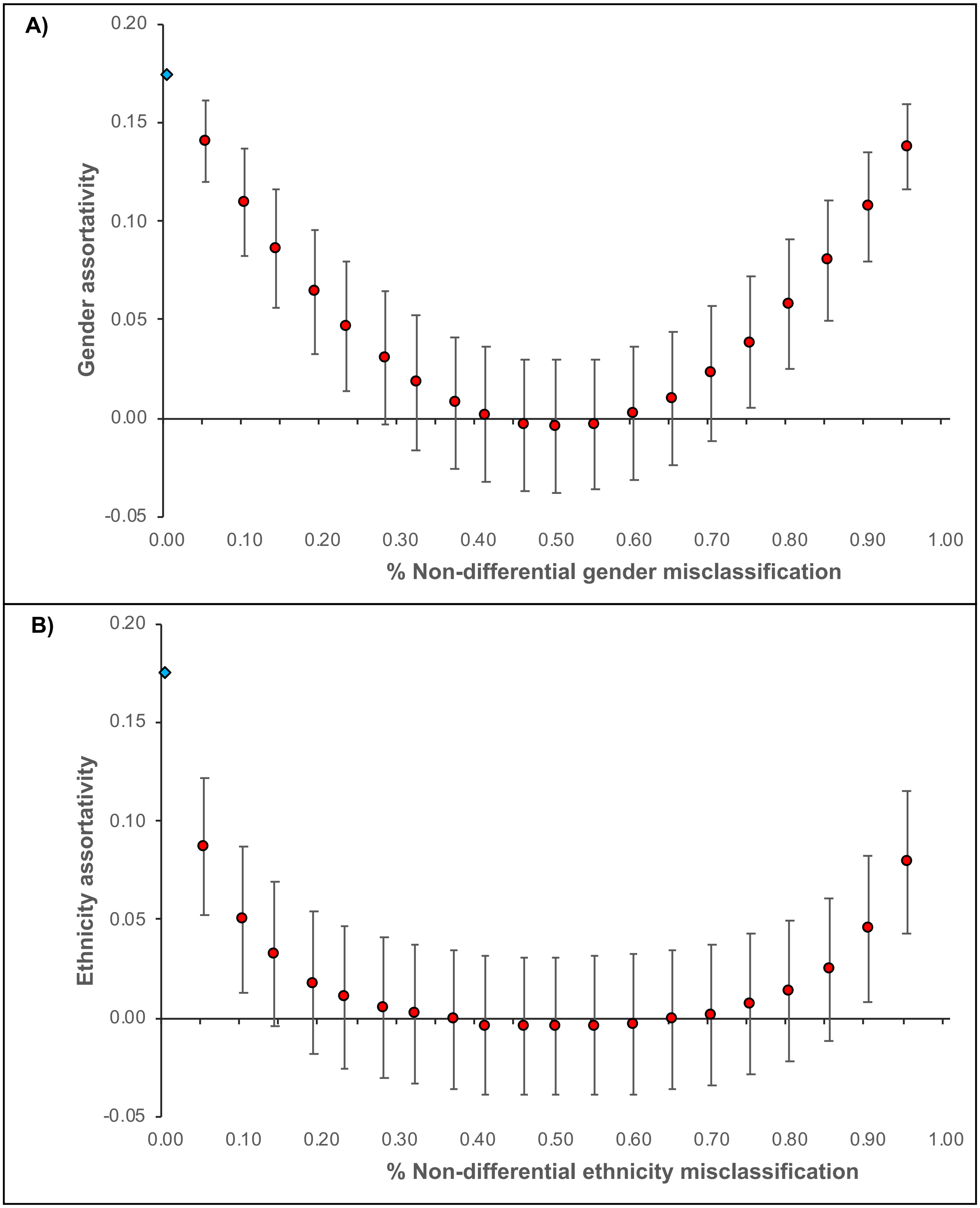


**Figure 4.9** Sensitivity analysis to assess impact of differential misclassification of alcohol use in the social network (2013), i.e., potential underreporting of alcohol use by all study participants who reported being nondrinkers at baseline (N=268 (64% of those with alcohol use data) and study participants under the legal drinking age (21 years; N=245 (91% of non-drinkers)). Each red dot or orange triangle and whiskers represent the mean (SD) attribute assortativity for 10,000 random samples of a given percent of differential misclassification. The blue dotted line represents the approximate expected prevalence of alcohol use in the study's undergraduate population, given previous research.

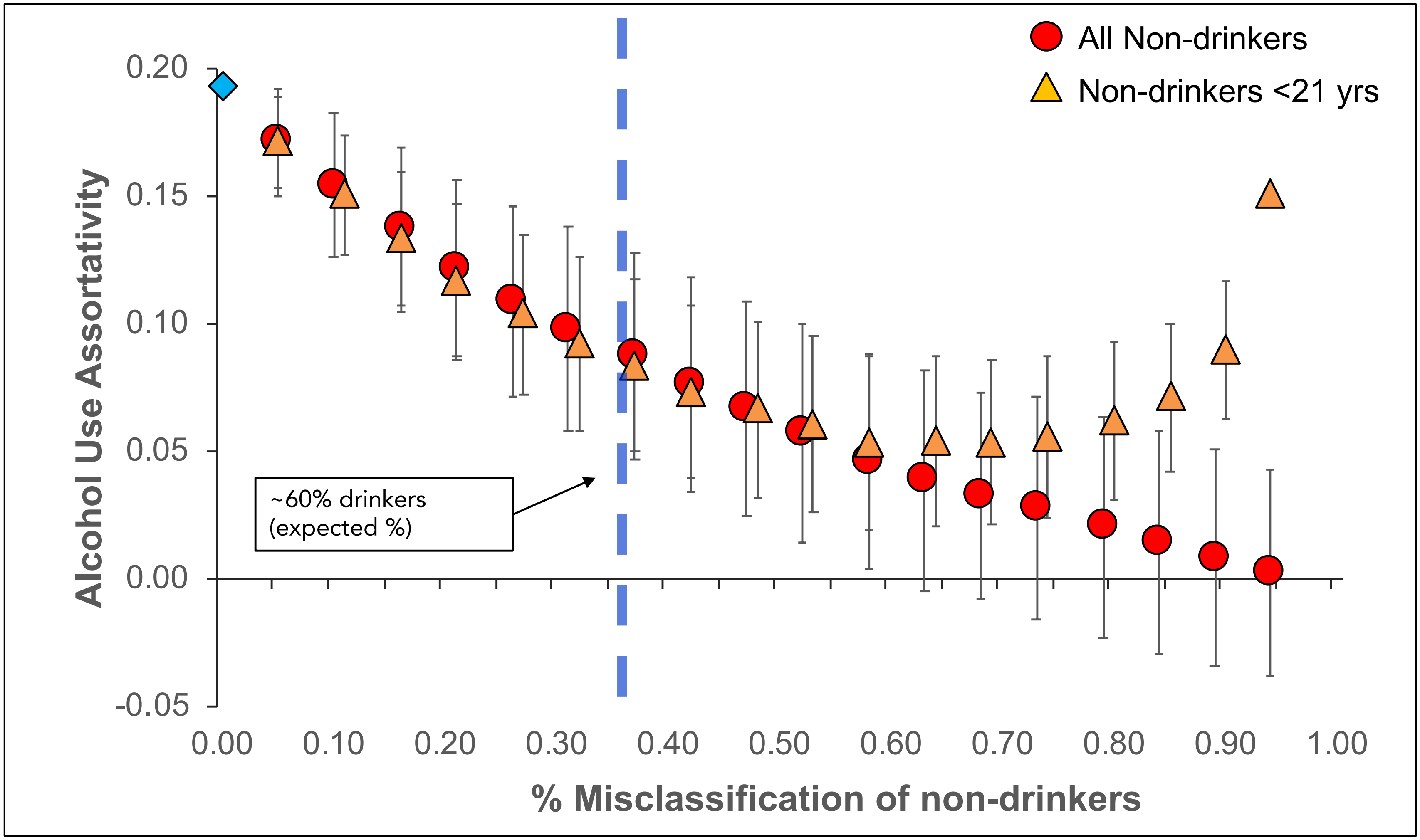


**Figure 4.10** Uncertainty quantification for alcohol use assortativity on the social network (2013). Each line segment represents the 95% confidence interval (CI) calculated from 10,000 resamples. Values without Cis resulted from undefined assortivities at a given level of jackknifing.

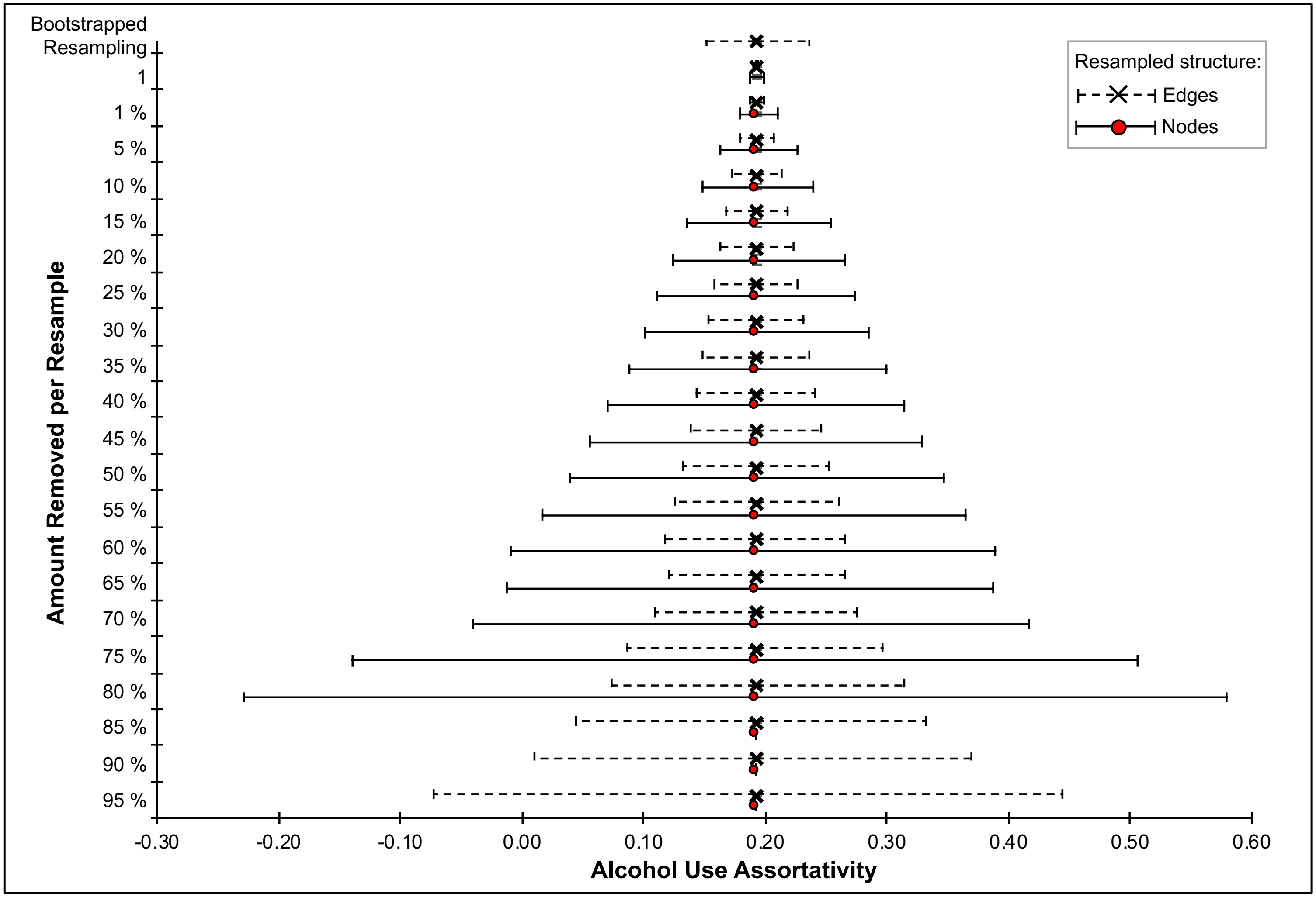


**Figure 4.11** Uncertainty quantification for ethnicity assortativity on the social network (2013). Each line segment represents the 95% confidence interval (CI) calculated from 10,000 resamples. Values without Cis resulted from undefined assortivities at a given level of jackknifing.

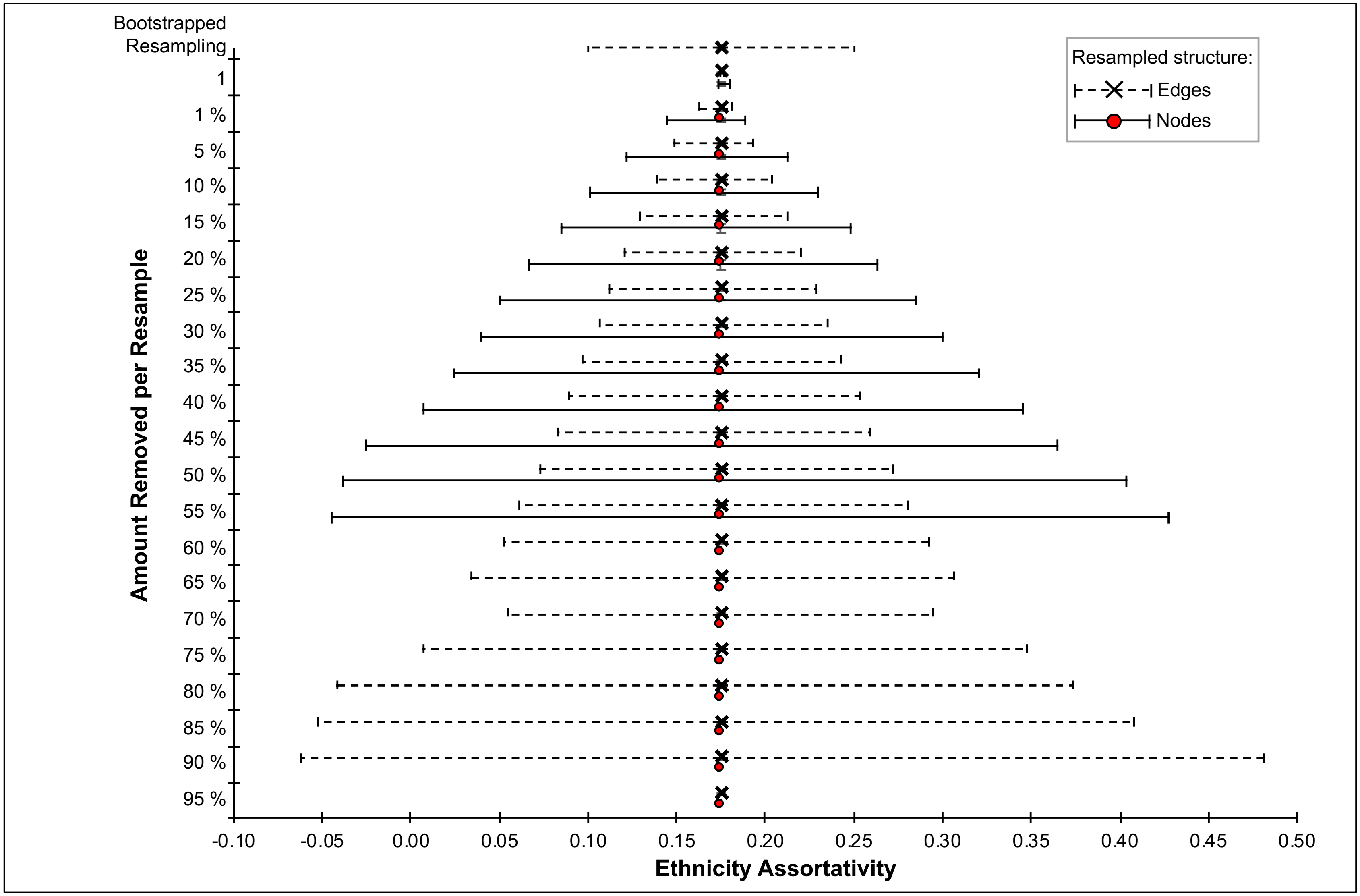


**Figure 4.12** Uncertainty quantification for gender assortativity on the social network (2013). Each line segment represents the 95% confidence interval (CI) calculated from 10,000 resamples. Values without Cis resulted from undefined assortivities at a given level of jackknifing.

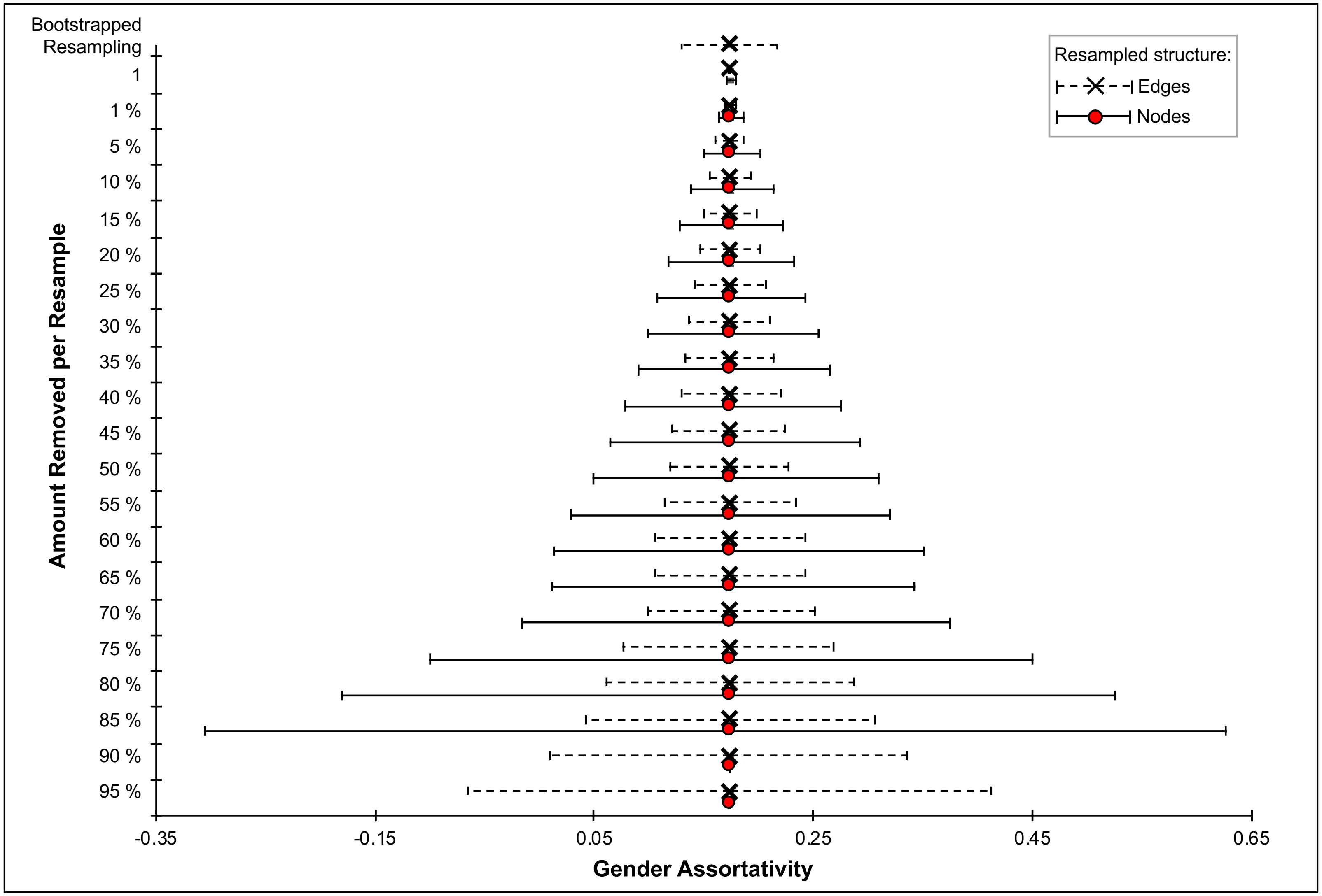


**Figure 4.13** Hypothesis testing results of node null model randomizations compared to the observed attribute assortativity on the social network (2013). Each panel shows the distribution of a given attribute's assortativities (A: alcohol use; B: ethnicity; C: gender) observed on the 10,000 resamples of the alcohol use node null model compared to the observed attribute assortativity on the measured network (red dotted line).

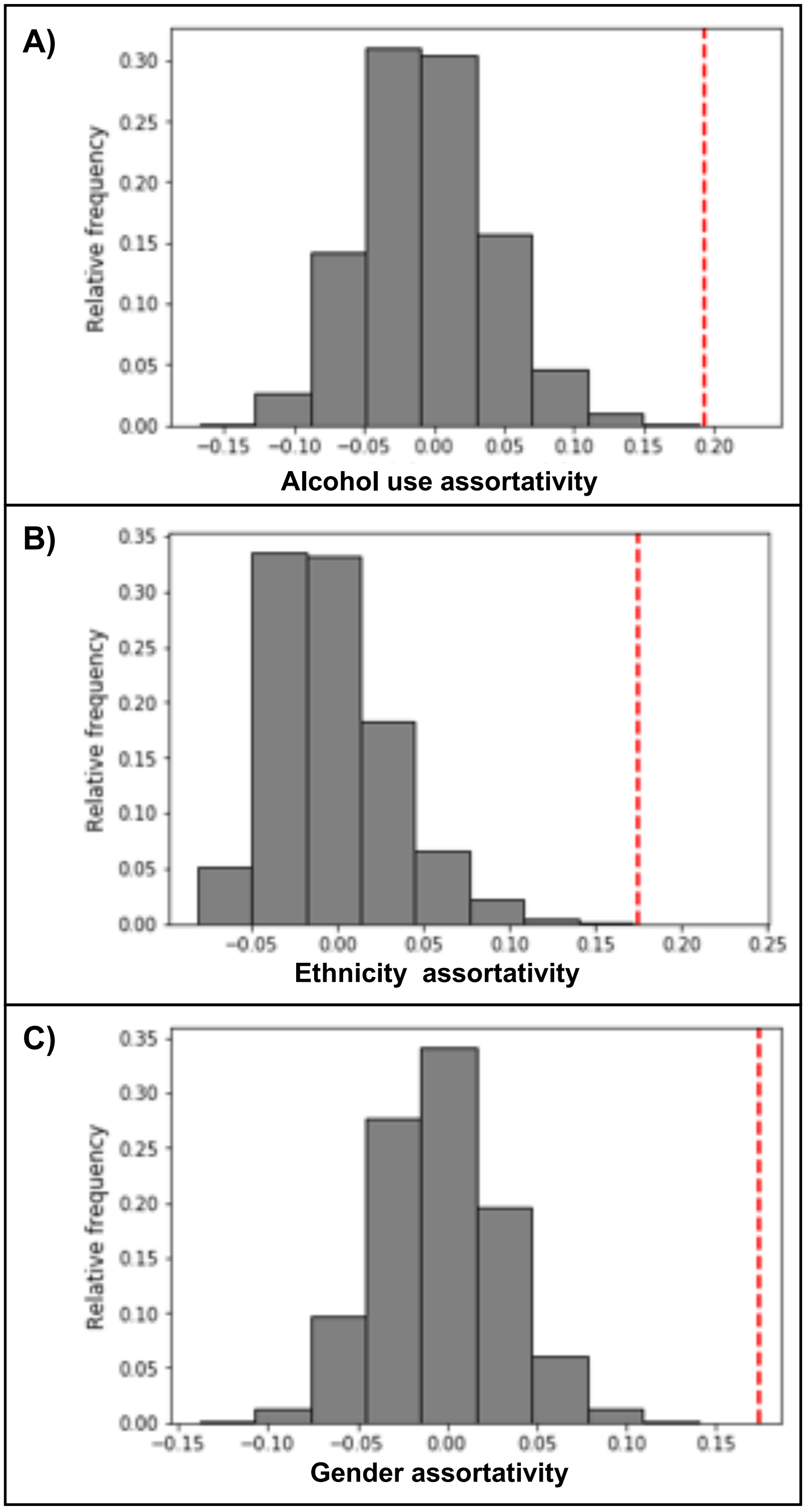
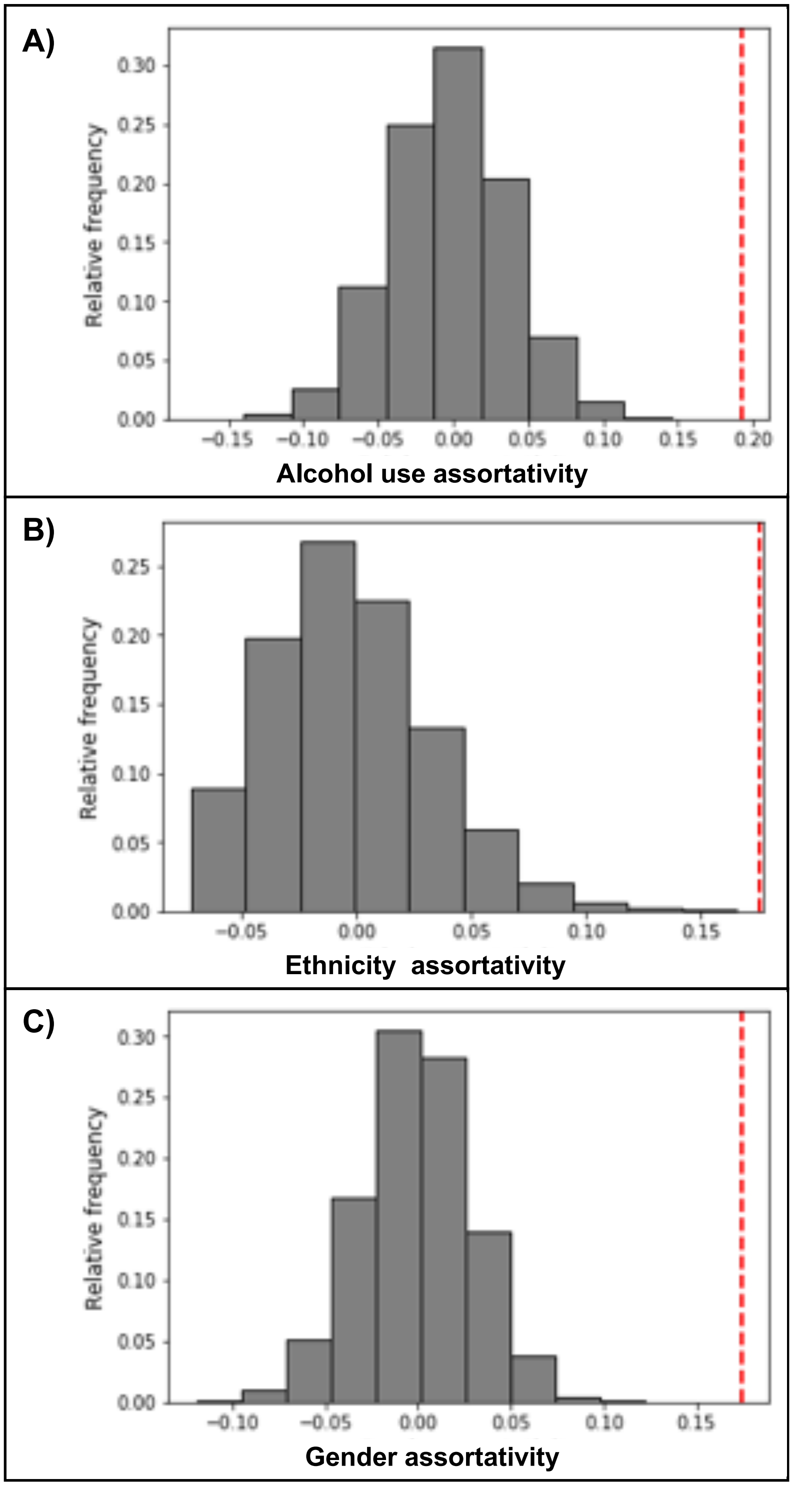


**Figure 4.14** Hypothesis testing results of edge null model randomizations compared to the observed attribute assortativity on the social network (2013). Each panel shows the distribution of a given attribute's assortativities (A: alcohol use; B: ethnicity; C: gender) observed on the 10,000 resamples of the alcohol use node null model compared to the observed attribute assortativity on the measured network (red dotted line).



# Chapter 5
# Conclusion

In the words of A.R.G. Owen, mathematician, geneticist, and parapsychology researcher, "[we] do not pretend to have answered [all] of these questions finally or even adequately in the present study. However, [we] feel that [we] have done something serviceable merely in assembling and discussing them…" (290)

A fundamental question that all social network research should address is: *how well does this data and model capture the human behavior of interest?* The overall aims of this dissertation were to increase understanding of: 1) how social structures and contact patterns shape alcohol consumption and use in undergraduate students; 2) drawbacks and data quality issues related to different methods of measuring social contact networks; and 3) how to articulate and quantify network uncertainty and hypothesis testing for trait assortativity.

Using data and social networks from eX-FLU study—a large, two-year social network study of university students, this dissertation applied existing social network analyses to college drinking and then defined and developed a number of novel methods for social network analysis in public health research including a direct network comparison statistic (specifically, an adapted Kappa coefficient), multiple sensitivity analyses and uncertainty quantification methods, and null models for hypothesis testing of trait assortativity.

In Chapter 2, we applied social network analysis methods to two networks of undergraduate students, investigating social network correlates of alcohol use and consumption volume. We identified numerous consistent associations between alcohol use and social position in this population. Specifically, network position, network alcohol exposures, and relationship strength were associated with individual alcohol use, suggesting complex relationships between drinking and network topology, as well as proximity to alcohol use. Overall, this chapter adds to the body of evidence of significant relationships between network structure, social position, and alcohol consumption. There appear to be complex interactions between social standing, demographics, individual motivations, and alcohol use; these associations, as well as their directionality, warrant further study. Notably, we identified numerous network structures and properties that present potential barriers to on-campus alcohol-abstinence interventions. However, additional studies are warranted to disentangle the complex network-alcohol associations we identified. In particular, we would like to investigate possible confounding of these relationships by age and/or Resident Advisorship by adapting and applying methods used by Block and Grund (59) to address multi-dimensional homophily.

Chapter 3 characterized and compared two set of longitudinal social networks, one collected with a Bluetooth-based app, iEpi, and the other using self-reported contacts. Networks were compared across and within each measurement method, using overall network structure, dyad, and node characteristics. Generally, we found few similarities between networks observed with iEpi and reported contacts, suggesting that neither empirical network measurement method are complete representations of the underlying "true" social network. In light of these findings, researchers should carefully consider how to best measure networks in terms of their research questions. We plan on extending this project in a number of ways, including: 1) a thorough investigation of longitudinal stability in the eX-FLU networks, in both the reported networks (2 years) and iEpi networks (1 year, and 2) comparisons of the infectious disease transmission patterns modeled on reported versus iEpi networks.

Given that current methods and instruments for measuring social networks may not capture their "true" underlying structure, in Chapter 4 we outlined methods to describe potential bias in observed attribute assortativity related to measurement error and nonresponse, as well as to quantify uncertainty and incorporate null model-based hypothesis testing into assortativity analyses. In so doing, we found that attribute assortativity biases and uncertainty may be predictable, given an *a priori* defined amount and type of data error. Overall, these methods provide a framework for attribute assortativity uncertainty analyses, as well as hypothesis testing, on sociocentric networks, particularly in contact network studies. We plan on extending these uncertainty methods to account for respondent-driven sampling, as well as to incorporate multiple types of data uncertainty and error in a one integrated method for uncertainty quantification. Additionally, applying the methods we developed to other empirical and theoretical networks (e.g., Add Health (128), small-world networks) may further clarify the nature of social network uncertainty.

Overall, this dissertation described and extensively explored social networks of undergraduate students. We investigated the relationships between a risky health behavior of public health importance and network features, as well as how network analysis results using observed networks are reliant on the network measurement method and the types and amounts of data uncertainty and error present. These projects generated new results and insights into alcohol use and social networks in a college setting, compared empirical social network observations between a traditional and novel instrument, and developed a suite of analytical social network tools. Importantly, the novel methods we have defined and implemented provide a framework with which to evaluate network uncertainty, robustness, and hypotheses.

# References

1. Aiello AE. Invited Commentary: Evolution of Social Networks, Health, and the Role of Epidemiology. *Am. J. Epidemiol.* 2017;185(11):1089–1092.

2. Snijders T, Steglich C, Schweinberger M. Modeling the Coevolution of Networks and Behavior. In: *Longitudinal Models in the Behavioral and Related Sciences*. Routledge; 2017:41–71.

3. Veenstra R, Dijkstra JK, Steglich C, et al. Network-Behavior Dynamics. *J. Res. Adolesc.* 2013;23(3):399–412.

4. Hayashi MAL, Eisenberg MC. Effects of adaptive protective behavior on the dynamics of sexually transmitted infections. *J. Theor. Biol.* 2016;388:119–30.

5. Reifman A, Watson WK, McCourt A. Social networks and college drinking: probing processes of social influence and selection. *Personal. Soc. Psychol. Bull.* 2006;32(6):820–832.

6. Meese N, Swart J, Vidgen R, et al. Addressing Data Collection Problems in Web-Mediated Surveys. In: *Volume 3: 30th Computers and Information in Engineering Conference, Parts A and B*. ASME; 2010:449–456.

7. Steglich C, Snijders TAB, Pearson M. 8. Dynamic Networks and Behavior: Separating Selection from Influence. *Sociol. Methodol.* 2010;40(1):329–393.

8. Funk S, Salathé M, Jansen VAA. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface*. 2010;7(50):1247–56.

9. Danon L, Ford AP, House T, et al. Networks and the epidemiology of infectious disease. *Interdiscip. Perspect. Infect. Dis.* 2011;2011:284909.

10. Granovetter M. The strength of weak ties. *Am. J. Sociol.* 1973;78(6):1360–1380.

11. Read JM, Eames KTD, Edmunds WJ. Dynamic social networks and the implications for the spread of infectious disease. *J. R. Soc. Interface*. 2008;5(26):1001–7.

12. Bansal S, Grenfell BT, Meyers LA. When individual behaviour matters: homogeneous and network models in epidemiology. *J R Soc Interface*. 2007;4(16):879–891.

13. Read JM, Edmunds WJ, Riley S, et al. Close encounters of the infectious kind: methods to measure social mixing behaviour. *Epidemiol. Infect.* 2012;140(12):2117–30.

14. Ferguson N. Capturing human behaviour. *Nature*. 2007;446(7137):733.

15. Salathé M, Jones JH. Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Comput. Biol.* 2010;6(4):e1000736.

16. Lloyd-Smith J, Schreiber S, Kopp P, et al. Superspreading and the effect of individual variation on disease emergence. *Nature*. 2005;438:355–359.

17. Newman MEJ. Assortative Mixing in Networks. *Phys. Rev. Lett.* 2002;89(20):208701.

18. Rushmore J, Caillaud D, Matamba L, et al. Social network analysis of wild chimpanzees provides insights for predicting infectious disease risk. *J. Anim. Ecol.* 2013;82(5):976–986.

19. Salathé M, Jones JH. Dynamics and Control of Diseases in Networks with Community Structure. *PLoS Comput. Biol.* 2010;6(4):e1000736.

20. Cohen S, Doyle WJ, Skoner DP, et al. Social ties and susceptibility to the common cold. *JAMA*. 1997;277(24):1940–1944.

21. Cohen S, Doyle WJ, Turner R, et al. Sociability and susceptibility to the common cold. *Psychol. Sci.* 2003;14(5):389–395.

22. Centers for Disease Control and Prevention (CDC) RJ, Brantley TR, Asfaw MM, et al. Notes from the field: measles outbreak among members of a religious community - Brooklyn, New York, March-June 2013. *MMWR. Morb. Mortal. Wkly. Rep.* 2013;62(36):752–3.

23. Hopkins Tanne J. New York City mayor declares measles public health emergency. *BMJ*. 2019;365:l1724.

24. Brotherton JML. The remarkable impact of bivalent HPV vaccine in Scotland. *BMJ*. 2019;365:l1375.

25. Aledort JE, Lurie N, Wasserman J, et al. Non-pharmaceutical public health interventions for pandemic influenza: an evaluation of the evidence base. *BMC Public Health*. 2007;7:208.

26. Jefferson T, Del Mar C, Dooley L, et al. Physical interventions to interrupt or reduce the spread of respiratory viruses: systematic review. *Bmj*. 2009;339:b3675.

27. Ferguson NM, Cummings DAT, Fraser C, et al. Strategies for mitigating an influenza pandemic. *Nature*. 2006;442(7101):448–452.

28. Bell D, Nicoll A, Fukuda K, et al. Non-pharmaceutical interventions for pandemic influenza, national and community measures. *Emerg. Infect. Dis.* 2006;12(1):88–94.

29. Deutsch AR, Steinley D, Slutske WS. The role of gender and friends' gender on peer

socialization of adolescent drinking: a prospective multilevel social network analysis. *J. Youth Adolesc.* 2014;43(9):1421–35.

30. Gaughan M. The gender structure of adolescent peer influence on drinking. *J. Health Soc. Behav.* 2006;47(1):47–61.

31. Clark AE, Lohéac Y. "It wasn't me, it was them!" Social influence in risky behavior by adolescents. *J. Health Econ.* 2007;26(4):763–784.

32. Jeon KC, Goodson P. US adolescents' friendship networks and health risk behaviors: a systematic review of studies using social network analysis and Add Health data. *PeerJ.* 2015;3:e1052.

33. Cintrón-Arias A, Sánchez F, Wang X, et al. The Role of Nonlinear Relapse on Contagion Amongst Drinking Communities. *Math. Stat. Estim. Approaches Epidemiol.* 2009;343–360.

34. Lau-Barraco C, Linden AN. Drinking buddies: Who are they and when do they matter? *Addict. Res. Theory*. 2014;22(1):57–67.

35. Douglas K, Collins J, Warren C, et al. Results from the 1995 national college health risk behavior survey. *J. Am.* 1997;46(2):55–66.

36. Kuntsche E, Knibbe R, Gmel G, et al. Why do young people drink? A review of drinking motives. *Clin. Psychol. Rev.* 2005;25(7):841–61.

37. Rosenquist JN, Murabito J, Fowler JH, et al. The spread of alcohol consumption behavior in a large social network. *Ann. Intern. Med.* 2010;152:426–433.

38. Mundt M. Social network analysis of peer effects on binge drinking among US adolescents. Berlin, Heidelberg, Heidelberg: Springer Berlin Heidelberg; 2013.

39. Mrug S, Borch C, Cillessen AHN. Other-sex friendships in late adolescence: Risky associations for substance use and sexual debut? *J. Youth Adolesc.* 2011;40(7):875–888.

40. Ah D Von, Ebert S, Ngamvitroj A, et al. Predictors of health behaviours in college students. *J. Adv. Nurs.* 2004;48(5):463–474.

41. Brady SS, Morrell HER, Song A V, et al. Longitudinal study of adolescents' attempts to promote and deter friends' smoking behavior. *J. Adolesc. Health*. 2013;53(6):772–7.

42. Christakis N, Fowler J. The collective dynamics of smoking in a large social network. *N. Engl. J. Med.* 2008;358:2249–2258.

43. Duncan G, Boisjoly J, Kremer M. Peer effects in drug use and sex among college students. *J. Abnorm. Child Psychol.* 2005;33(3):375–385.

44. Halebsky MA. Adolescent alcohol and substance abuse: parent and peer effects.

*Adolescence*. 1987;22(88):961–967.

45. Leibsohn J. The Relationship Between Drug and Alcohol Use and Peer Group Associations of College Freshmen as They Transition from High School. *J. Drug Educ.* 1994;24(3):177–192.

46. Barclay KJ, Edling C, Rydgren J. Peer clustering of exercise and eating behaviours among young adults in Sweden: a cross-sectional study of egocentric network data. *BMC Public Health*. 2013;13(1):784.

47. Barnett NP, Ott MQ, Rogers ML, et al. Peer associations for substance use and exercise in a college student social network. *Health Psychol.* 2014;33(10):1134–42.

48. Yakusheva O, Kapinos K, Weiss M. Peer effects and the Freshman 15: Evidence from a natural experiment. *Econ. Hum. Biol.* 2011;9(2):119–132.

49. Richmond TK, Milliren C, Walls CE, et al. School Social Capital and Body Mass Index in the National Longitudinal Study of Adolescent Health. *J. Sch. Health*. 2014;84(12):759–768.

50. Christakis N, Fowler J. The spread of obesity in a large social network over 32 years. *N. Engl. J. Med.* 2007;357:370–379.

51. Moore S, Daniel M, Paquet C, et al. Association of individual network social capital with abdominal adiposity, overweight and obesity. *J. Public Health (Bangkok).* 2008;31(1):175–183.

52. Lotar-Rihtarić M, Kamenov Ž. Susceptibility to peer pressure and attachment to friends. *Psihologija*. 46(2):111–126.

53. Adams GR. Physical Attractiveness, Personality, and Social Reactions to Peer Pressure. *J. Psychol.* 1977;96(2):287–296.

54. French DA, Teymuroglu Z, Lewis TJ, et al. An integro-differential equation model for the spread of alcohol abuse. *J. Integr. Equations Appl.* 2010;22:443–464.

55. Allen JP, Chango J, Szwedo D, et al. Predictors of susceptibility to peer influence regarding substance use in adolescence. *Child Dev.* 2012;83(1):337–50.

56. Edmunds WJ, Kafatos G, Wallinga J, et al. Mixing patterns and the spread of close-contact infectious diseases. *Emerg. Themes Epidemiol.* 2006;3:10.

57. Newman MEJ. Mixing patterns in networks. *Phys. Rev. E*. 2003;67(2):26126.

58. Brechwald WA, Prinstein MJ. Beyond Homophily: A Decade of Advances in Understanding Peer Influence Processes. *J. Res. Adolesc.* 2011;21(1):166–179.

59. Block P, Grund T. Multidimensional homophily in friendship networks. *Netw. Sci.*

2014;2(2):189–212.

60. Guo G, Li Y, Wang H, et al. Peer Influence, Genetic Propensity, and Binge Drinking: A Natural Experiment and a Replication. *AJS.* 2015;121(3):914–54.

61. Ali MM, Amialchuk A, Nikaj S. Alcohol consumption and social network ties among adolescents: evidence from Add Health. *Addict. Behav.* 2014;39(5):918–922.

62. Hingson RW, Zha W, Weitzman ER. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18-24, 1998-2005. *J. Stud. Alcohol Drugs. Suppl.* 2009;(16):12–20.

63. O'Malley PPM, Johnston LLD. Epidemiology of alcohol and other drug use among American college students. *J. Stud. Alcohol. Suppl.* 2002;(14):23–39.

64. McCarty C, Killworth PD, Rennell J. Impact of methods for reducing respondent burden on personal network structural measures. *Soc. Networks*. 2007;29(2):300–315.

65. Room R, Babor T, Rehm J. Alcohol and public health. *Lancet*. 2005;(365):519–30.

66. Grace T. Health problems of college students. *J Am Coll Heal.* 1997;45(6):243–250.

67. Rehm J, Mathers C, Popova S, et al. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet.* 2009;373(9682):2223–2233.

68. White A, Hingson R. The burden of alcohol use: excessive alcohol consumption and related consequences among college students. *Alcohol Res. Curr. Rev.* 2014;35(2):201–218.

69. McCambridge J, McAlaney J, Rowe R. Adult consequences of late adolescent alcohol consumption: A systematic review of cohort studies. *PLoS Med.* 2011;8(2).

70. Hingson RW. Focus On: College Drinking and Related Problems Magnitude and Prevention of College Drinking and Related Problems. *Alcohol Res. Heal.* 2010;33(1–2):45–54.

71. HINGSON R, Assailly J, Williams A. Underage drinking: frequency, consequences, and interventions. *Traffic Inj. Prev.* 2004;5(3):228–236.

72. Johnston LD, O'Malley PM, Miech RA, et al. Monitoring the future national survey results on drug use, 1975–2015. 2016.

73. Slutske WS, Hunt-Carter EE, Nabors-Oberg RE, et al. Do college students drink more than their non-college-attending peers? Evidence from a population-based longitudinal female twin study. *J. Abnorm. Psychol.* 2004;113(4):530–540.

74. National Institute on Alcohol Abuse and Alcoholism (NIAAA). Drinking Levels Defined.

2016;(https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking)

75. Nelson JP, Young DJ. Effects of youth, price, and audience size on alcohol advertising in magazines. *Health Econ.* 2008;17(4):551–6.

76. Knight JR, Wechsler H, Kuo M, et al. Alcohol abuse and dependence among U.S. college students. *J. Stud. Alcohol*. 2002;63(3):263–70.

77. Mallett KA, Varvil-Weld L, Borsari B, et al. An Update of Research Examining College Student Alcohol-Related Consequences: New Perspectives and Implications for Interventions. *Alcohol. Clin. Exp. Res.* 2013;37(5):709–716.

78. Wechsler H, Nelson TF. What We Have Learned From the Harvard School of Public Health College Alcohol Study: Focusing Attention on College Student Alcohol Consumption and the Environmental Conditions That Promote It. *J. Stud. Alcohol Drugs*. 2008;69(4):481–490.

79. Wechsler H, Lee JE, Nelson TF, et al. Underage college students' drinking behavior, access to alcohol, and the influence of deterrence policies. Findings from the Harvard School of Public Health College Alcohol Study. *J Am. Coll. Health*. 2002;50(5):223–36.

80. Wechsler H, Molnar BE, Davenport AE, et al. College alcohol use: a full or empty glass? *J Am. Coll. Health*. 1999;47(6):247–52.

81. White AM, Kraus CL, Swartzwelder HS. Many college freshmen drink at levels far beyond the binge threshold. *Alcohol. Clin. Exp. Res.* 2006;30(6):1006–1010.

82. Read JP, Beattie M, Chamberlain R, et al. Beyond the &quot; Binge &quot; threshold: Heavy drinking patterns and their association with alcohol involvement indices in college students. *Addict. Behav.* 2008;33:225–234.

83. Perkins H. Surveying the damage: A review of research on consequences of alcohol misuse in college populations. *J. Stud. Alcohol Drugs*. 2002;(14):91.

84. Presley C a, Meilman PW, Leichliter JS. College factors that influence drinking. *J. Stud. Alcohol. Suppl.* 2002;(14):82–90.

85. Hingson RW, Heeren T, Zakocs RC, et al. Magnitude of alcohol-related mortality and morbidity among U.S. College students ages 18-24. *J. Stud. Alcohol*. 2002;63(2):136–144.

86. Cooper ML. Alcohol use and risky sexual behavior among college students and youth: evaluating the evidence. *J. Stud. Alcohol, Suppl.* 2002;(s14):101–117.

87. Mallett KA, Varvil-Weld L, Turrisi R, et al. An examination of college students' willingness to experience consequences as a unique predictor of alcohol problems. *Psychol. Addict. Behav.* 2011;25(1):41–7.

88. Patrick ME, Maggs JL. Short-term changes in plans to drink and importance of positive and negative alcohol consequences. *J. Adolesc.* 2008;31(3):307–321.

89. Mallett KA, Bachrach RL, Turrisi R. Are all negative consequences truly negative? Assessing variations among college students' perceptions of alcohol related consequences. *Addict. Behav.* 2008;33(10):1375–1381.

90. Nguyen N, Walters ST, Wyatt TM, et al. Do College Drinkers Learn from Their Mistakes? Effects of Recent Alcohol-Related Consequences on Planned Protective Drinking Strategies among College Freshmen. *Subst. Use Misuse*. 2013;48(14):1463–1468.

91. Borsari B, Carey KBK. Peer influences on college drinking: A review of the research. *J. Subst. Abuse*. 2001;13(4):391–424.

92. Bishop DI, Weisgram ES, Holleque KM, et al. Identity development and alcohol consumption: Current and retrospective self-reports by college students. *J. Adolesc.* 2005;28(4):523–533.

93. McCabe SE, Boyd CJ, Cranford JA, et al. Alcohol involvement and participation in residential learning communities among first-year college students. *J. Stud. Alcohol Drugs*. 2007;68(5):722–726.

94. Capone C, Wood MD, Borsari B, et al. Fraternity and Sorority Involvement, Social Influences, and Alcohol Use Among College Students: A Prospective Examination. *Psychol. Addict. Behav.* 2007;21(3):316–327.

95. Grekin ER, Sher KJ. Alcohol dependence symptoms among college freshmen: Prevalence, stability, and person-environment interactions. *Exp. Clin. Psychopharmacol.* 2006;14(3):329–338.

96. Hartzler B, Fromme K. Heavy episodic drinking and college entrance. *J. Drug Educ.* 2003;33(3):259–274.

97. Boyd CJ, McCabe SE, Morales M. College students' alcohol use: a critical review. *Annu. Rev. Nurs. Res.* 2005;23:179–211.

98. Weitzman ERE, Nelson TFT, Wechsler H. Taking up binge drinking in college: The influences of person, social group, and environment. *J. Adolesc. Heal.* 2003;32(1):26–35.

99. White HR, McMorris BJ, Catalano RF, et al. Increases in alcohol and marijuana use during the transition out of high school into emerging adulthood: The effects of leaving home, going to college, and high school protective factors. *J. Stud. Alcohol*. 2006;67(6):810–22.

100. Read JP, Wood MD, Davidoff OJ, et al. Making the Transition from High School to College: The Role of Alcohol-Related Social Influence Factors in Students' Drinking. *Subst. Abus.* 23(1):53–65.

101. Reboussin BA, Song E-Y, Wolfson M. Social influences on the clustering of underage risky drinking and its consequences in communities. *J. Stud. Alcohol Drugs*. 2012;73(6):890–8.

102. Ladd BO, Hallgren KA, Greenfield BL. The Moderating Effects of Drinking Motives on Social Network and Individual Drinking in College Students. *Alcohol. Exp. Res.* 2011;35(6):30A-30A.

103. Valente TW. Social network influences on adolescent substance use: An introduction. *Connections*. 2003;25(2):11–16.

104. Wood M, Read J, Palfai T, et al. Social influence processes and college student drinking: the mediational role of alcohol outcome expectancies. *J. Stud. Alcohol*. 2001;61(1):32–43.

105. Villarosa M, Kison S, Madson M, et al. Everyone else is doing it: examining the role of peer influence on the relationship between social anxiety and alcohol use behaviours. *Addict. Res. Theory*. 2016;24(2):124–134.

106. Simons-Morton B, Haynie D, Liu D, et al. The effect of residence, school status, work status, and social influence on the prevalence of alcohol use among emerging adults. *J Stud Alcohol Drugs*. 2016;77(1):121–132.

107. Borsari B, Carey KB. How the quality of peer relationships influences college alcohol use. *Drug Alcohol Rev.* 2006;25(4):361–370.

108. Fletcher JM. Peer influences on adolescent alcohol consumption: Evidence using an instrumental variables/fixed effect approach. *J. Popul. Econ.* 2012;25(4):1265–1286.

109. Kandel DB. On processes of peer influences in adolescent drug use: a developmental perspective. *Adv. Alcohol Subst. Abuse*. 4(3–4):139–63.

110. Kandel DB. Drug and Drinking Behavior Among Youth. *Annu. Rev. Sociol.* 1980;6:235–285.

111. Perkins HW. The Contextual Effect of Secular Norms on Religiosity as Moderator of Student Alcohol and other Drug Use. *Res. Soc. Sci. Study Relig.* 1994;6:187–208.

112. Orford J, Krishnan M, Balaam M, et al. University student drinking: The role of motivational and social factors. *Drugs Educ. Prev. Policy*. 2004;11(5):407–421.

113. Astudillo M, Connor J, Roiblatt RE, et al. Influence from friends to drink more or drink less: A cross-national comparison. *Addict. Behav.* 2013;38(11):2675–2682.

114. Jaccard J, Blanton H, Dodge T. Peer influences on risk behavior: An analysis of the effects of a close friend. *Dev. Psychol.* 2005;41(1):135–147.

115. Glassman T, Braun RE, Dodd V, et al. Using the theory of planned behavior to explain the drinking motivations of social, high-risk, and extreme drinkers on game day. *J.*

*Community Health*. 2010;35(2):172–81.

116. Baer JS. Student factors: Understanding individual variation in college drinking. *J. Stud. Alcohol*. 2002;63(Suppl. 14):40–53.

117. Larimer ME, Anderson BK, Baer JS, et al. An individual in context: Predictors of alcohol use and drinking problems among Greek and residence hall students. *J. Subst. Abuse*. 2000;11(1):53–68.

118. Gonzalez VM, Skewes MC. Solitary heavy drinking, social relationships, and negative mood regulation in college drinkers. *Addict. Res. Theory*. 2013;21(4):285–294.

119. Demant J, Järvinen M. Social capital as norms and resources: Focus groups discussing alcohol. *Addict. Res. Theory*. 2011;19(2):91–101.

120. Weitzman E, Nelson T, Wechsler H. Taking up binge drinking in college: The influences of person, social group, and environment. *J. Adolesc. Heal.* 2003;32(1):26–35.

121. Ham LS, Hope DA. College students and problematic drinking: A review of the literature. *Clin. Psychol. Rev.* 2003;23:719–759.

122. Dowdall G, Wechsler H. Studying college alcohol use: widening the lens, sharpening the focus. *J. Stud. Alcohol, Suppl.* 2002;14:14–22.

123. Leichliter JS, Meilman PW, Presley CA, et al. Alcohol Use and Related Consequences Among Students With Varying Levels of Involvement in College Athletics. *J. Am. Coll. Heal.* 1998;46(6):257–262.

124. Martens MP, Dams-O'Connor K, Beck NC. A systematic review of college student-athlete drinking: Prevalence rates, sport-related factors, and interventions. *J. Subst. Abuse Treat.* 2006;31(3):305–316.

125. Mizruchi MS, Marquis C. Egocentric, sociocentric, or dyadic?Identifying the appropriate level of analysis in the study of organizational networks. *Soc. Networks*. 2006;28(3):187–208.

126. Wasserman S, Faust K. Social network analysis: Methods and applications. New York, NY, NY: Cambridge University Press, Publishers; 1994.

127. Deutsch AR, Slutske WS, Richmond-Rakerd LS, et al. Causal influence of age at first drink on alcohol involvement in adulthood and its moderation by familial context. *J. Stud. Alcohol Drugs*. 2013;74(5):703–13.

128. Harris KM, Halpern CT, Whitsel E, et al. The National Longitudinal Study of Adolescent to Adult Health: Research Design. 2009;

129. Cruz JE, Emery RE, Turkheimer E. Peer network drinking predicts increased alcohol use from adolescence to early adulthood after controlling for genetic and shared

environmental selection. *Dev. Psychol.* 2012;48(5):1390–402.

130. Fujimoto K, Valente TW. Social network influences on adolescent substance use: Disentangling structural equivalence from cohesion. *Soc. Sci. Med.* 2012;74(12):1952–1960.

131. Ali MM, Dwyer DS. Social network effects in alcohol consumption among adolescents. *Addict. Behav.* 2010;35(4):337–342.

132. Mundt MP, Eaten DK, Kann L, et al. The impact of peer social networks on adolescent alcohol use initiation. *Acad. Pediatr.* 2006;11(5):414–21.

133. Fujimoto K, Valente TWTW, Alexander C, et al. Decomposing the components of friendship and friends' influence on adolescent drinking and smoking. *J. Adolesc. Health.* 2012;51(2):136–43.

134. Fujimoto K, Valente TW. Alcohol peer influence of participating in organized school activities: a network approach. *Health Psychol.* 2013;32(10):1084–92.

135. Phua J. The influence of peer norms and popularity on smoking and drinking behavior among college fraternity members: A social network analysis. *Soc. Influ.* 2011;6(3):153–168.

136. Dumas TM, Wells S, Flynn A, et al. The Influence of Status on Group Drinking by Young Adults: A Survey of Natural Drinking Groups on Their Way To and From Bars. *Alcohol. Clin. Exp. Res.* 2013;38(4):1100–7.

137. Dumas T, Graham K, Bernards S, et al. Drinking to reach the top: Young adults' drinking patterns as a predictor of status within natural drinking groups. *Addict. Behav.* 2014;39(10):1510–1515.

138. Dumas TM, Graham K, Maxwell-Smith MA, et al. Being cool is risky business: Young men's within-peer-group status, heavy alcohol consumption and aggression in bars. *Addict. Res. Theory*. 2015;23(3):213–222.

139. Eisenberg D, Golberstein E, Whitlock J. Peer effects on risky behaviors: New evidence from college roommate assignments. *J. Health Econ.* 2014;33:126–138.

140. Barnett NP, Ott M, Clark M. The relevance of network prominence and reciprocity of relationships for alcohol use and alcohol-related problems in a college residence hall network. *Psychol. Addict. Behav.* 2014;28(4):980–9.

141. Barnett N, Ott M, Rogers M. Peer associations for substance use and exercise in a college student social network. *Heal. Psychol.* 2014;33(10):1134–1142.

142. Lorant V, Nicaise P. Binge drinking at University: a social network study in Belgium. *Health Promot. Int.* 2015;30(3):675–83.

143. VanderWeele T, Ogburn E, Tchetgen E. Why and when “flawed” social network analyses still yield valid tests of no contagion. *Stat Polit. Policy*. 2012;3(1):2151–1050.

144. Chung KKS, Hossain L, Davis J. Exploring sociocentric and egocentric approaches for social network analysis. *Int. Conf. Knowl. Manag. Asia Pacific, Novemb. 27–29, Victoria Univ. Wellington, New Zeal.* 2005;1–8.

145. Bell DC, Atkinson JS, Carlson JW. Centrality measures for disease transmission networks. *Soc. Networks*. 1999;21(1):1–21.

146. Bonacich P. Power and centrality: A family of measures. *Am. J. Sociol.* 1987;92(5):1170–1182.

147. Imamverdiyev Y, Asl H, Janani N, et al. A longitudinal study on degree centrality changes in a group of students. *4th Int. Conf. Appl. Inf. Commun. Technol.* 2010;

148. Watts DJ, Strogatz SH. Collective dynamics of “small-world” networks. *Nature*. 1998;393(6684):440–2.

149. Beauchamp M. An improved index of centrality. *Behav. Sci.* 1965;10(2):161–3.

150. Freeman LC. Segregation in social networks. *Sociol. Methods Res.* 1978;6(4):411–429.

151. Newman M. Networks: an introduction. Oxford University Press; 2010.

152. Fisher D, Silk M, Franks D. The perceived assortativity of social networks: methodological problems and solutions. In: Missaoui R, Abdessalem T, Latapy T, eds. *Trends in Social Network Analysis. Lecture Notes in Social Networks*. Springer, Cham; 2017:1–19.

153. Newman MEJ, Park J. Why social networks are different from other types of networks. *Phys. Rev. E*. 2003;68(3):036122.

154. Mossong J, Hens N, Jit M, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med.* 2008;5(3):e74.

155. Franks D, Ruxton G, James R. Sampling animal association networks with the gambit of the group. *Behav. Ecol. Sociobiol.* 2010;64(3):493–503.

156. Knoke CD, Yang S. Social Network Analysis Basic Methods for Analyzing Networks Title: Social Network Analysis Chapter Title: “Basic Methods for Analyzing Networks.” 2008 92–118 p.

157. Hawe P, Webster C, Shiell A. A glossary of terms for navigating the field of social network analysis. *J. Epidemiol. Community Health*. 2004;58(12):971–5.

158. Browne K. Snowball sampling: using social networks to research non-heterosexual women. *Int. J. Soc. Res. Methodol.* 2005;8(1):47–60.

159. Johnston LG, Sabin K. Sampling Hard-to-Reach Populations with Respondent Driven Sampling. *Methodol. Innov. Online*. 2010;5(2):38.1-48.

160. Kolaczyk E. Statistical analysis of network data: Methods and Models. Springer; 2009.

161. Mills HL, Johnson S, Hickman M, et al. Errors in reported degrees and respondent driven sampling: implications for bias. *Drug Alcohol Depend.* 2014;142:120–6.

162. Heckathorn DD, Cameron CJ. Network Sampling: From Snowball and Multiplicity to Respondent-Driven Sampling. *Annu. Rev. Sociol.* 2017;43(1):101–119.

163. Brañas-Garza P, Jiménez N, Ponti G. A guided tour to (real-life) social network elicitation. *J. Behav. Econ. Policy*. 2017;1(1):1–10.

164. Smith J a., Moody J. Structural effects of network sampling coverage I: Nodes missing at random. *Soc. Networks*. 2013;35(4):652–668.

165. Koskinen JH, Robins GL, Wang P, et al. Bayesian analysis for partially observed network data, missing ties, attributes and actors. *Soc. Networks*. 2013;35(4):514–527.

166. Huisman M, Steglich C. Treatment of non-response in longitudinal network studies. *Soc. Networks*. 2008;30(4):297–308.

167. Marchette DJ, Hohman EL. Utilizing covariates in partially observed networks. In: *2015 18th International Conference on Information Fusion, Fusion 2015*. Institute of Electrical and Electronics Engineers Inc.; 2015:166–172.

168. Robins G, Pattison P, Woolcock J. Missing data in networks: exponential random graph (p∗) models for networks with non-respondents. *Soc. Networks*. 2004;26(3):257–283.

169. McCaw JM, Forbes K, Nathan PM, et al. Comparison of three methods for ascertainment of contact information relevant to respiratory pathogen transmission in encounter networks. *BMC Infect. Dis.* 2010;10(1):166.

170. Kossinets G. Effects of missing data in social networks. *Soc. Networks*. 2006;28(3):247–268.

171. Hoang T Van, Coletti P, Melegaro A, et al. A systematic review of social contact surveys to inform transmission models of close contact infections. *bioRxiv*. 2018;292235.

172. Frantz TL, Cataldo M, Carley KM. Robustness of centrality measures under uncertainty: Examining the role of network topology. *Comput. Math. Organ. Theory*. 2009;15(4):303–328.

173. Leung K, Jit M, Lau EHY, et al. Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Sci. Rep.* 2017;7(1):7974.

174. Aiello AEA, Simanek AM, Eisenberg MMC, et al. Design and methods of a social

network isolation study for reducing respiratory infection transmission: The eX-FLU cluster randomized trial. *Epidemics*. 2016;15:38–55.

175. Barabási A-L, Albert R. Emergence of scaling in random networks. *Science (80-. ).* 1999;286(5439):509–512.

176. Brañas-Garza P, Cobo-Reyes R, Espinosa MP, et al. Altruism and social integration. *Games Econ. Behav.* 2010;69(2):249–257.

177. Marin A. Are respondents more likely to list alters with certain characteristics? Implications for name generator data. *Soc. Networks*. 2004;26(4):289–307.

178. Brewer DD. Forgetting in the recall-based elicitation of personal and social networks. *Soc. Networks*. 2000;22(1):29–43.

179. Farine DR, Whitehead H. Constructing, conducting and interpreting animal social network analysis. *J. Anim. Ecol.* 2015;84(5):1144–1163.

180. Farine D. Estimating uncertainty and reliability of social network data using Bayesian inference. *R. Soc. open sci*. 2015;2:150367.

181. Stork D, Richards WD. Nonrespondents in Communication Network Studies. *Gr. Organ. Manag.* 1992;17(2):193–209.

182. Eames K, Bansal S, Frost S, et al. Six challenges in measuring contact networks for use in modelling. *Epidemics*. 2014;10:72–77.

183. Feld SL, Carter WC. Detecting measurement bias in respondent reports of personal networks. *Soc. Networks*. 2002;24(4):365–383.

184. Mastrandrea R, Fournet J, Barrat A. Contact Patterns in a High School: A Comparison between Data Collected Using Wearable Sensors, Contact Diaries and Friendship Surveys. *PLoS One*. 2015;10(9):e0136497.

185. Leecaster M, Toth DJ a., Pettey WBP, et al. Estimates of Social Contact in a Middle School Based on Self-Report and Wireless Sensor Data. *PLoS One*. 2016;11(4):e0153690.

186. Barrat a., Cattuto C, Tozzi a. E, et al. Measuring contact patterns with wearable sensors: Methods, data characteristics and applications to data-driven simulations of infectious diseases. *Clin. Microbiol. Infect.* 2014;20(1):10–16.

187. Smieszek T, Barclay VC, Seeni I, et al. How should social mixing be measured: comparing web-based survey and sensor-based methods. *BMC Infect. Dis.* 2014;14:136.

188. Isella L, others. Close Encounters in a Pediatric Ward: Measuring Face-to-Face Proximity and Mixing Patterns with Wearable Sensors. *PLoS One*. 2011;6(2):e17144.

189. Voirin N, Payet C, Barrat A, et al. Combining high-resolution contact data with

virological data to investigate influenza transmission in a tertiary care hospital. *Infect. Control Hosp. Epidemiol.* 2015;36(3):254–60.

190. Kazandjieva MA, Lee JW, Salathé M, et al. Experiences in measuring a human contact network for epidemiology research. In: *Proceedings of the 6th Workshop on Hot Topics in Embedded Networked Sensors - HotEmNets '10*. New York, New York, USA: ACM Press; 2010:1.

191. Funk S, Bansal S, Bauch CT, et al. Nine challenges in incorporating the dynamics of behaviour in infectious diseases models. *Epidemics*. 2015;10:21–5.

192. Aggarwal CC, Abdelzaher T. Integrating sensors and social networks. *Soc. Netw. Data Anal.* 2011;

193. Klasnja P, Consolvo S, Choudhury T, et al. Exploring Privacy Concerns about Personal Sensing. In: H. T, M. B, A. F, et al., eds. *Pervasive Computing. Pervasive 2009. Lecture Notes in Computer Science, vol 5538*. Berlin, Heidelberg: Springer; 2009

194. Smieszek T, Castell S, Barrat A, et al. Contact diaries versus wearable proximity sensors in measuring contact patterns at a conference: method comparison and participants' attitudes. *BMC Infect. Dis.* 2016;16(1):341.

195. Mastrandrea R, Barrat A, Barrat A, et al. How to Estimate Epidemic Risk from Incomplete Contact Diaries Data? *PLOS Comput. Biol.* 2016;12(6):e1005002.

196. Fournet J, Barrat A, Eames K, et al. Epidemic risk from friendship network data: an equivalence with a non-uniform sampling of contact networks. *Sci. Rep.* 2016;6:24593.

197. Wang L, Chu C, Yang G, et al. Transmission characteristics of different students during a school outbreak of (H1N1) pdm09 influenza in China, 2009. *Sci. Rep.* 2014;4:5982.

198. Hens N, Goeyvaerts N, Aerts M, et al. Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. *BMC Infect. Dis.* 2009;9:5.

199. Edmunds WJ, O'callaghan CJ, Nokes DJ. Who mixes with whom? A method to determine the contact patterns of adults that may lead to the spread of airborne infections. *Proc. R. Soc. B Biol. Sci.* 1997;264(1384):949–957.

200. Walker WE, Harremoës P, Rotmans J, et al. Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support. *Integr. Assess.* 2003;4(1):5–17.

201. Albert R, Jeong H, Barabási A-L. Error and attack tolerance of complex networks. *Nature*. 2000;406:378–382.

202. Carley KM, Lee J-S, Krackhardt D. Destabilizing Networks. *Connections*. 2001;23(3):79–92.

203. Bliss CA, Danforth CM, Dodds PS. Estimation of global network statistics from incomplete data. *PLoS One*. 2014;9(10):e108471.

204. Hoffman FO, Hammonds JS. Propagation of Uncertainty in Risk Assessments: The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability. *Risk Anal.* 1994;14(5):707–712.

205. Shizuka D, Farine DR. Measuring the robustness of network community structure using assortativity. *Anim. Behav.* 2016;112:237–246.

206. Borgatti SP, Carley KM, Krackhardt D. On the robustness of centrality measures under conditions of imperfect data. *Soc. Networks*. 2006;28(2):124–136.

207. Bernard HR, Johnsen EC, Killworth PD, et al. Comparing four different methods for measuring personal social networks. *Soc. Networks*. 1990;12(3):179–215.

208. Carley K. Dynamic network analysis. In: Brieger R, Carley K, eds. *Summary of the NRC workshop on Social Network Modeling and Analysis*. National Research Council; 2003

209. Kossinets G. Effects of missing data in social networks. *Soc. Networks*. 2006;28(3):247–268.

210. Croft DP, Madden JR, Franks DW, et al. Hypothesis testing in animal social networks. *Trends Ecol. Evol.* 2011;26(10):502–507.

211. Jennison KM. The short-term effects and unintended long-term consequences of binge drinking in college: a 10-year follow-up study. *Am. J. Drug Alcohol Abuse*. 2004;30(3):659–84.

212. National Institute on Alcohol Abuse and Alcoholism. A call to action: Changing the culture of drinking at U.S. Colleges. Final report of the Task Force on College Drinking. NIH Pub. No. 02-5010. Rockville, MD: 2002.

213. L.D. Johnston, P.M. O'Malley, J.G. Bachman, J.E. Schulenberg RAM. Monitoring the future National Survey Results on Drug Use, 1975–2015, Volume 2: College Students and Adults Ages 19-55. 2016.

214. Borsari B, Murphy JG, Barnett NP. Predictors of alcohol use during the first year of college: implications for prevention. *Addict. Behav.* 2007;32(10):2062–86.

215. Baer JS. Student factors: Understanding individual variation in college drinking. *J. Stud. Alcohol*. 2002;63(SUPPL. 14):40–53.

216. Arnett JJ. The Developmental Context of Substance use in Emerging Adulthood. *J. Drug Issues*. 2005;35(2):235–254.

217. Demant J, Järvinen M. Social capital as norms and resources: Focus groups discussing alcohol. *Addict. Res. Theory*. 2010;19(2):91–101.

218. Oliphant TE. Python for scientific computing. *Comput. Sci. Eng.* 2007;9(3):10–20.

219. Hesterberg T, Monaghan S, Moore D. Bootstrap methods and permutation tests. In: Company WHF and, ed. *THE PRACTICE OF BUSINESS STATISTICS*. New York, New York, USA:

220. Fisher N, Statistics PH-AJ of, 1990 undefined. On bootstrap hypothesis testing. *Wiley Online Libr.*

221. Hall P, Hart JD. Bootstrap Test for Difference between Means in Nonparametric Regression. *J. Am. Stat. Assoc.* 1990;85(412):1039–1049.

222. Balsa AI, Homer JF, French MT, et al. Alcohol use and popularity: Social payoffs from conforming to peers' behavior. *J. Res. Adolesc.* 2011;21(3):559–568.

223. Leonard KE, Mudar P. Peer and partner drinking and the transition to marriage: a longitudinal examination of selection and influence processes. *Psychol. Addict. Behav.* 2003;17(2):115–25.

224. Giese H, Stok FM, Renner B. The Role of Friendship Reciprocity in University Freshmen's Alcohol Consumption. *Appl. Psychol. Heal. Well-Being*. 2017;9(2):228–241.

225. Centola D. The Spread of Behavior in an Online Social Network Experiment. *Science (80-. ).* 2010;329(5996):1194–1197.

226. Dierker L, Stolar M, Lloyd-Richardson E, et al. Tobacco, alcohol, and marijuana use among first-year U.S. college students: a time series analysis. *Subst. Use Misuse*. 2008;43(5):680–99.

227. Young CM, DiBello AM, Steers M-LN, et al. I like people who drink like me: Perceived appeal as a function of drinking status. *Addict. Behav.* 2016;53:125–131.

228. Meisel MK, DiBello AM, Balestrieri SG, et al. An event- and network-level analysis of college students' maximum drinking day. *Addict. Behav.* 2018;79:189–194.

229. Young CM, DiBello AM, Traylor ZK, et al. A Longitudinal Examination of the Associations Between Shyness, Drinking Motives, Alcohol Use, and Alcohol-Related Problems. *Alcohol. Clin. Exp. Res.* 2015;39(9):1749–1755.

230. Schry AR, White SW. Understanding the relationship between social anxiety and alcohol use in college students: A meta-analysis. *Addict. Behav.* 2013;38:2690–2706.

231. Buckner JD, Eggleston AM, Schmidt NB. Social Anxiety and Problematic Alcohol Consumption: The Mediating Role of Drinking Motives and Situations. *Behav. Ther.* 2006;37(4):381–391.

232. Gilles DM, Turk CL, Fresco DM. Social anxiety, alcohol expectancies, and self-efficacy as predictors of heavy drinking in college students. *Addict. Behav.* 2006;31(3):388–398.

233. Bruch MA, Rivet KM, Heimberg RG, et al. Shyness, alcohol expectancies, and drinking behavior: Replication and extension of a suppressor effect. *Pers. Individ. Dif.* 1997;22(2):193–200.

234. University Health Service. Summary of the Results of the National College Health Assessment University of Michigan, Ann Arbor Campus. 2014.

235. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction*. 2003;98(s2):1–12.

236. Adams CE, Nagoshi CT. Changes over one semester in drinking game playing and alcohol use and problems in a college student sample. *Subst. Abus.* 1999;20(2):97–106.

237. Hoeppner BB, Barnett NP, Jackson KM, et al. Daily college student drinking patterns across the first year of college. *J. Stud. Alcohol Drugs*. 2012;73(4):613.

238. Beets MW, Flay BRBR, Vuchinich S, et al. Longitudinal patterns of binge drinking among first year college students with a history of tobacco use. *Drug Alcohol Depend.* 2009;103(1–2):1–8.

239. Boca F Del, Darkes J. Up close and personal: temporal variability in the drinking of individual college students during their first year. *J. Consult. Clin. Psychol.* 2004;72(2):155–64.

240. White AM, Kraus CL, McCracken LA, et al. Do College Students Drink More Than They Think? Use of a Free-Pour Paradigm to Determine How College Students Define Standard Drinks. *Alcohol. Clin. Exp. Res.* 2003;27(11):1750–1756.

241. White AM, Kraus CL, Flom JD, et al. College students lack knowledge of standard drink volumes: implications for definitions of risky drinking based on survey data. *Alcohol. Clin. Exp. Res.* 2005;29(4):631–8.

242. Mercken L, Snijders TAB, Steglich C, et al. Smoking-based selection and influence in gender-segregated friendship networks: A social network analysis of adolescent smoking. *Addiction*. 2010;105(7):1280–1289.

243. Kiuru N, Burk WJ, Laursen B, et al. Pressure to drink but not to smoke: disentangling selection and socialization in adolescent peer networks and peer groups. *J. Adolesc.* 2010;33(6):801–812.

244. Rinker DV, Krieger H, Neighbors C. Social Network Factors and Addictive Behaviors Among College Students. *Curr. Addict. Reports*. 2016;356–367.

245. Valente T. Social networks and health: Models, methods, and applications. New York, NY: Oxford University Press; 2010.

246. Fenichel EP, Castillo-Chavez C, Ceddia MG, et al. Adaptive human behavior in epidemiological models. *Proc. Natl. Acad. Sci.* 2011;108(15):6306–6311.

247. Wallinga J, Edmunds WJ, Kretzschmar M. Perspective: human contact patterns and the spread of airborne infectious diseases. *Trends Microbiol.* 1999;7(9):372–7.

248. Mikolajczyk RT, Kretzschmar M. Collecting social contact data in the context of disease transmission: Prospective and retrospective study designs. *Soc. Networks*. 2008;30(2):127–135.

249. Mayer A, Puller SL. The old boy (and girl) network: Social network formation on university campuses. *J. Public Econ.* 2008;92(1):329–347.

250. Eames KTD, Tilston NL, White PJ, et al. The impact of illness and the impact of school closure on social contact patterns. *Health Technol. Assess.* 2010;14(34):267–312.

251. Cornwell B. Age Trends in Daily Social Contact Patterns. *Res. Aging*. 2015;33(5):598–631.

252. Wallinga J, Teunis P, Kretzschmar M. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *Am. J. Epidemiol.* 2006;164(10):936–44.

253. Yang H-H, Wei C-K, Yang H-J. The dynamic friendship networks and health behavior of university students in Southern Taiwan. *Taiwan J. Public Heal.* 2008;27(4):341–348.

254. Bansal S, Read J, Pourbohloul B, et al. The dynamic nature of contact networks in infectious disease epidemiology. *J. Biol. Dyn.* 2010;4(5):478–489.

255. Stomakhin A, Short MB, Bertozzi AL. Reconstruction of missing data in social networks based on temporal patterns of interactions. *Inverse Probl.* 2011;27(11):115013.

256. Feld SL. Structural embeddedness and stability of interpersonal relations. *Soc. Networks*. 1997;19(1):91–95.

257. Felmlee DH. No Couple Is an Island: A Social Network Perspective on Dyadic Stability. *Soc. Forces*. 2001;79(4):1259–1287.

258. Leventhal E a, Hansell S, Diefenbach M, et al. Negative affect and self-report of physical symptoms: two longitudinal studies of older adults. *Health Psychol.* 1996;15(3):193–199.

259. Perkins SE, Cagnacci F, Stradiotto A, et al. Comparison of social networks derived from ecological data: implications for inferring infectious disease dynamics. *J. Anim. Ecol.* 2009;78(5):1015–22.

260. Smieszek T, Burri EU, Scherzinger R, et al. Collecting close-contact social mixing data with contact diaries: reporting errors and biases. *Epidemiol. Infect.* 2012;140(4):744–52.

261. Potter GE, Handcock MS, Longini IM, et al. ESTIMATING WITHIN-HOUSEHOLD CONTACT NETWORKS FROM EGOCENTRIC DATA. *Ann. Appl. Stat.* 2011;5(3):1816–1838.

262. Potter GE, Smieszek T, Sailer K. Modelling workplace contact networks: the effects of organizational structure, architecture, and reporting errors on epidemic predictions. 2013;

263. Salathé M, Kazandjieva M, Lee JW, et al. A High-Resolution Human Contact Network for Infectious Disease Transmission. *Proc. Natl. Acad. Sci.* 2010;107(51):22020–22025.

264. Paik A, Sanchagrin K. Social Isolation in America: An Artifact. *SSRN Electron. J.* 2012;

265. Eagle DE, Proeschold-Bell RJ. Methodological considerations in the use of name generators and interpreters. *Soc. Networks*. 2015;40:75–83.

266. Hashemian M, Knowles D, Calver J, et al. iEpi: an end to end solution for collecting, conditioning and utilizing epidemiologically relevant data. In: *Proceedings of the 2nd ACM international workshop on Pervasive Wireless Healthcare*. 2012:3–8.

267. Girvan M, Newman MEJ. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U. S. A.* 2002;99(12):7821–7826.

268. Knight WR. A Computer Method for Calculating Kendall's Tau with Ungrouped Data. *J. Am. Stat. Assoc.* 1966;61(314, Part 1):436–439.

269. Arndt S, Turvey C, Andreasen NC. Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation. *J. Psychiatr. Res.* 33(2):97–104.

270. McHugh ML. Interrater reliability: the kappa statistic. *Biochem. medica*. 2012;22(3):276–82.

271. Kim KJ, Shin D-H. An acceptance model for smart watches. *Internet Res.* 2015;25(4):527–541.

272. Kalantari M. Consumers adoption of wearable technologies: literature review, synthesis, and future research agenda. *Int. J. Technol. Mark.* 2017;12(1):1.

273. Costenbader E, Valente TW. The stability of centrality measures when networks are sampled. *Soc. Networks*. 2003;25(4):283–307.

274. Marsden P V. The reliability of network density and composition measures. *Soc. Networks*. 1993;15(4):399–421.

275. Zemljič B, Hlebec V. Reliability of measures of centrality and prominence. *Soc. Networks*. 2005;27(1):73–88.

276. on PS-P of the 2nd S workshop, 2008 undefined. Social network analysis of uncertain networks. *pdfs.semanticscholar.org*.

277. Hanneman R, Riddle M. Introduction to Social Network Methods. Riverside, CA: University of California; 2005.

278. Sundaresan S, Fischhoff I, Behaviour JD-A, et al. Avoiding spurious findings of

nonrandom social structure in association data. *Elsevier*.

279. Krause S, Mattner L, James R, et al. Social network analysis and valid Markov chain Monte Carlo tests of null models. *Springer*.

280. Dormann C. What's wrong with null models? 2017;(https://theoreticalecology.wordpress.com/2017/06/06/whats-wrong-with-null-models/). (Accessed June 15, 2018)

281. Bejder L, Fletcher D, Behaviour SB-A, et al. A method for testing association patterns of social animals. *Elsevier*.

282. Delaney J, Seeger J. Sensitivity analysis. In: Velentgas P, Dreyer N, Nourjah P, eds. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide.* Rockville (MD): Agency for Healthcare Research and Quality (US); 2013

283. Marsden P V. Network Data and Measurement. *Annu. Rev. Sociol.* 1990;16(1):435–463.

284. Mcknight P, McKnight K, Guilford SS-, et al. Missing data: A Gentle Introduction. New York, New York, USA: Guilford Press; 2007.

285. Hanhijärvi S, Garriga GC, Puolamäki K. Randomization Techniques for Graphs. In: *Proceedings of the 2009 SIAM International Conference on Data Mining*. Philadelphia, PA: Society for Industrial and Applied Mathematics; 2009:780–791.

286. Lusseau D, Whitehead H, Gero S. Incorporating uncertainty into the study of animal social networks. *Anim. Behav.* 2008;75(5):1809–1815.

287. Cohen R, Erez K, ben-Avraham D, et al. Resilience of the Internet to Random Breakdowns. *Phys. Rev. Lett.* 2000;85(21):4626–4628.

288. Robins G, Pattison P, Networks JW-S, et al. Missing data in networks: exponential random graph (p∗) models for networks with non-respondents. *Soc. Networks*. 2004;26(3):257–283.

289. Fu YC, Ho HC, Chen HM. Weak ties and contact initiation in everyday life: Exploring contextual variations from contact diaries. *Soc. Networks*. 2013;35(3):279–287.

290. Owen AR. Can we explain the poltergeist? New York, NY: Helix Press/Garrett Publications; 1964.

# Appendix

## A.1 eX-FLU Study weekly social contact survey

The dynamics of social contacts and networks of individuals varies over time due to changes in activities, classes, and relationships. We would like to have the most current information about the people enrolled in the study that you have had face-to-face contact with in the past week.

Below is a list of enrolled individuals you reported being in contact with in the past, as well as others who are enrolled in the study that you may or may not have had face-to-face contact with recently. Please look over this list and click the "add" button next to every person that you have had face-to-face contact from [Date] to [Date].

| **Past Social Contacts** | **Potential Social Contacts** |
|---|---|
| [first name] [last name] (email address)<br>[add] | [first name] [last name] (email address)<br>[add] |
| [first name] [last name] (email address)<br>[add] | [first name] [last name] (email address)<br>[add] |
| [If list is empty]<br>No enrolled individuals meet these criteria | [If list is empty]<br>No enrolled individuals meet these criteria |

If there is someone that you have had face-to-face contact with from [Date] to [Date] who was not listed above, please search for them by their email or full name below and then click the add button next to their name in the search results. Please note, only students who are enrolled in the study will show up in the search results.
_______________[search]

**[If person not found]**
The person you searched for was not found. You either misspelled their name or email or they are not enrolled in the study. You may try your search again or search for additional people you had face-to-face contact from [Date] to [Date] by returning to the search field above.

Below is a list of all the people you reported you've had contact with from [Date] to [Date]. If you wish to select additional individuals that you have had contact with, you may still do so by clicking the "add" button next to their name in the ‘Search Results’ table above or by returning to the search tool above.

| **People you have reported contact**<br>**From [Date] to [Date]** |
|---|
| [first name] [last name] (email address) **[remove]** |
| [first name] [last name] (email address) **[remove]** |

If you accidentally added someone you did not have face-to-face contact with and you wish to remove them from the list, please click on the “remove” button next to their name below. Your responses are confidential and will not be shared with any other individuals in the study. When you have finished selecting all the individuals you've had contact with from [Date] to [Date], please click the "next" button to continue the survey.

| **Next >>** |
|---|

**[if less than three contacts selected]**
**Would you like to select more social contacts?**
You have selected [0-2] contacts. Ideally we would like to collect information about your face-to-face interactions with at least three of your social contacts. If you are having trouble using the interface please e-mail [study staff email]. Otherwise, please select an option to continue:

I would like to select more contacts: [select more contacts]

I would like to continue the survey with [x] social contacts reported.

## A.2 eX-FLU Study Enrollment Survey (excerpt)

**In order to complete the enrollment process, you must complete the following enrollment survey.**

*Your responses are confidential. Please respond as accurately as possible. Thank you!*

***The next section focuses on your demographic information.***

**What is your gender?**

| | |
|---|---|
| | Male |
| | Female |
| | Don’t know |
| | decline to answer |

***The following questions ask about your ethnicity and race. All government funded projects are required to report according to these breakdown categories. We appreciate your willingness to work within these guidelines.***

**Which ethnicity best describes you?**

| | |
|---|---|
| | Hispanic or Latino |
| | Not Hispanic or not Latino |
| | Don’t know |
| | Decline to answer |

**Which race best describes you?**

| | |
|---|---|
| | White - a person having origins in any of the original peoples of Europe, North America, or the Middle East. |
| | Black or African American |
| | Asian - a person of the Far East, Southeast Asia, or the Indian subcontinent. |
| | Native Hawaiian or Other Pacific Islander |
| | American Indian or Alaskan Native - a person having origins in any of the original peoples of North America, and who maintains cultural identification through tribal affiliation or community recognition. |
| | Multi-Ethnic - a person having origins in more than one of the groups listed above. |
| | Don’t know |
| | Decline to answer |

**[If Multi-Ethnic] Please select the races to which you belong.**

| | |
|---|---|
| | White |
| | Black or African American |
| | Asian |
| | Native Hawaiian or Other Pacific Islander |
| | American Indian or Alaskan Native |
| | Don't know |
| | Decline to answer |

**What is your year of study, by number of years at the university?**

| | |
|---|---|
| | Freshman |
| | Sophomore |
| | Junior |
| | Senior |
| | Senior + |
| | Decline to answer |

**Are you a US citizen?**

| | |
|---|---|
| | Yes |
| | No |
| | Decline to answer |

## A.3 eX-FLU Study Baseline (Week 1) Survey (excerpt)

***The following questions pertain to your alcohol consumption.***

**Do you currently drink alcohol at least once a week?**

| | |
|---|---|
| | Yes |
| | No |
| | Don’t know |
| | Decline to answer |

**[If yes] During an average week, how much do you usually drink each day? (One drink is equal to a glass of wine, 12 ounce beer, or shot of hard liquor.)**

| | Nothing | 1 drink | 2 drinks | 3 drinks | 4 drinks | 5 drinks | 6 or more drinks | Don’t know | Decline to answer |
|---|---|---|---|---|---|---|---|---|---|
| **Monday** | | | | | | | | | |
| **Tuesday** | | | | | | | | | |
| **Wednesday** | | | | | | | | | |
| **Thursday** | | | | | | | | | |
| **Friday** | | | | | | | | | |
| **Saturday** | | | | | | | | | |
| **Sunday** | | | | | | | | | |