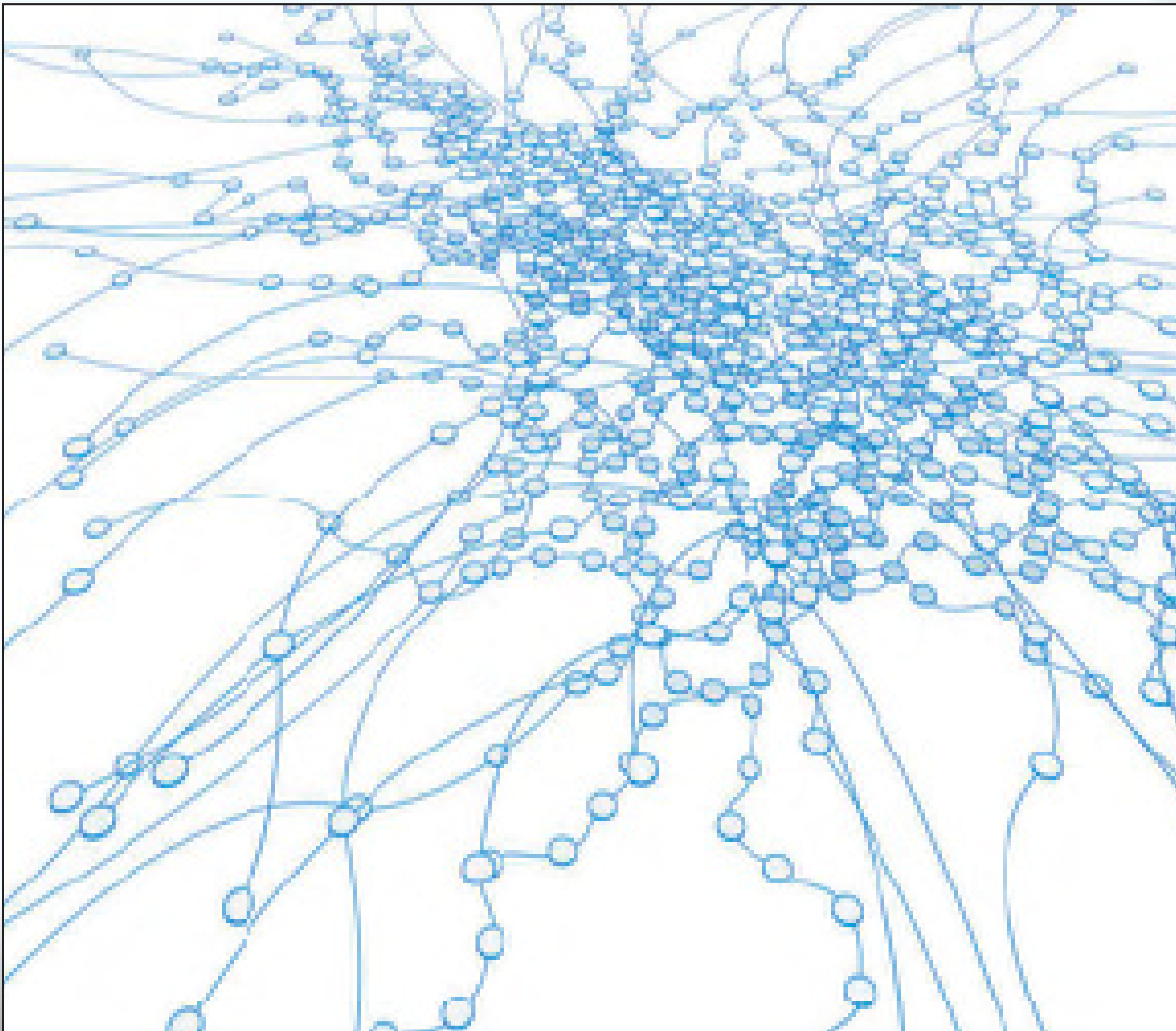
Connections

Volume 34, Issues 1&2 December 2014



Official Journal of the International Network for Social Network Analysis

INSNA
INTERNATIONAL NETWORK FOR SOCIAL NETWORK ANALYSIS

Connections publishes original empirical articles that apply social network analysis. We are particularly interested in research reports, articles demonstrating methodological innovations and novel research design. We also facilitate the dissemination of datasets and review scholarly books. Our scope spans many disciplines and domains including Communication, Anthropology, Sociology, Psychology, Organizational Behavior, Knowledge Management, Marketing, Social Psychology, Political Science, Public Health, Policy Studies, Medicine, Physics, Economics, Mathematics, and Computer Science. As the official journal for the International Network for Social Network Analysis (INSNA), the emphasis of the publication is to reflect the ever-growing and continually expanding community of scholars using network analytic techniques.

Background figure: Tighe, this issue pg. 36.

Connections
Manuscripts selected for publication undergo a double-blind peer-review process. See Manuscript Submissions for instructions and formatting details.

From the Editor

Welcome to a new issue of Connections.

I am pleased to report on a new team in charge of copyediting, layout, online presence and journal management. A number of changes on the style and presentation of the journal are planned over the next twelve months. From this issue onwards we introduce dedicated DOI numbers to each article (backdated to 2012). This will allow citing papers as soon as final proofs are accepted. By integrating this process to our review we anticipate to substantially reduce the gap from submission to an accessible online version of an article. We have expanded the DEN section and introduced a book review section with a view to evaluate the expanding literature of textbooks and monographs in the field. We also intend to retain a print run of the journal as we are committed to supply print issues for our institutional and library subscribers. From next year we are considering reducing the mailing costs by making the journal available at Sunbelt for all attending or to charge postage to those requiring a copy by post.

In this issue you can find articles on networks of the unemployed (Harris, et al.), bipartite flow centrality (Gerdes), hospital operating room ties (Tighe et al.) and mobile phone data collection (Comulada). We also publish four Data Exchange Network (DEN) articles on football club rivalries in Argentina (Bundio), terrorist networks in Mali (Walther and Christopoulos), collaboration networks in the 2013 Sunbelt (Pfeffer, Hollstein and Skvoretz) and the transactions in a local currency group in Portland (Collom). You will also find here a report on the 2014 Sunbelt keynote of Jeff Johnson (van Holt), two book reviews and a report from the last INSNA Business Meeting.

I take this opportunity to solicit good quality research papers that will be of interest to an interdisciplinary audience of network analysts. Papers are double blind reviewed and we aim to complete a first review within three months. I am particularly keen on papers with methodological or research design insights. With short papers with methods or results that have inter-disciplinary appeal being particularly welcome. I am also pleased to welcome a new associate editor with responsibility for DEN, Juergen Pfeffer. I would like to renew our invitation to share your codebook and data with the SNA community. DEN invites short articles that explain research design choices, potential limitations and originality in a dataset.

Representing the editorial team I would like to conclude by welcoming your feedback and invite you to a reception we plan to hold on Wednesday the 24th of June during the Brighton Sunbelt.

Dimitris Christopoulos
Editor, Connections
www.dimitriscc.wordpress.com

Connections

Employment Networks in a High-Unemployment Rural Area

Jenine K. Harris
Washington University in St. Louis
St. Louis, Missouri

Ellen Barnidge
Saint Louis University
St. Louis, Missouri

Elizabeth A. Baker
Saint Louis University
St. Louis, Missouri

Freda Motton
Saint Louis University
St. Louis, Missouri

Catherine Radvanyi
Saint Louis University
St. Louis, Missouri

Frank Rose
Saint Louis University
St. Louis, Missouri

Abstract
Higher rates of unemployment are found among African-American men in rural communities in the US. As part of a community-based participatory research project, we sought to identify characteristics of job-seeking networks of African-American and white employed and unemployed men and women in a rural community in Missouri. We collected cross-sectional quantitative and qualitative information about job-seeking networks through in-depth interviews with 39 local residents. Descriptive network measures were used to compare the gender, race, and employment status of the people comprising participant job-seeking networks. A novel network approach was used to simulate a whole network from individual networks depicting likely patterns of job-seeking relationships across the community. Unemployed participants had larger networks, with the exception of white women. Men had more racially homogenous networks than women; many networks had no racial diversity. Men had longer relationships than women, while women had stronger relationships. Employed participants had more linkages to alters with connections to community organizations than unemployed participants. Unemployed participants had many connections, but lacked connections to the right people and organizations to aid in their job search. Increasing employment opportunities in this community, and similar communities, will require effort from job-seekers and others to develop new relationships, programs, and policies.

Keywords: employment, social networks, rural, race

Authors
Jenine K. Harris, George Warren Brown School of Social Work, Washington University in St. Louis.

Elizabeth A. Baker, Saint Louis University College for Public Health & Social Justice.

Catherine Radvanyi, Saint Louis University College for Public Health & Social Justice.

Ellen Barnidge, Saint Louis University College for Public Health & Social Justice.

Freda Motton, Saint Louis University College for Public Health & Social Justice

Frank Rose, Saint Louis University College for Public Health & Social Justice.

Notes
This research was supported by a grant from the United States National Institutes of Health National Center on Minority Health and Health Disparities (NIH CMHD 5R24MD001590-06; EA Baker, PI). The authors would like to acknowledge the invaluable contributions of our participants and partners in Pemiscot County, Missouri.

Correspondence concerning this work should be addressed to Jenine K. Harris, George Warren Brown School of Social Work, Washington University in St. Louis, St. Louis, MO 63130, 314-935-3522, Email harrisj@wustl.edu

1. Introduction

Social ties and social capital have been identified as important in job-seeking (Granovetter, 1983; Lin, Cook, & Burt, 2001). The strength of an individual's social ties and the characteristics of the people they are connected to including race, gender, social class, and presence of social capital, are important to understanding the influence of these ties on employment. First, weak social ties (e.g., acquaintances and friends-of-friends) may be beneficial in job-seeking because they bring new information to a relationship. Stronger social connections tend to be among individuals who have similar characteristics, similar connections, and similar information. Weak ties tend to act as bridges between strongly connected groups which carry new information between the groups (Friedkin, 1980; Granovetter, 1983; Wegener, 1991).

Different demographic groups tend toward different types of ties in their personal social networks which may influence the effectiveness of social ties for influencing employment (McDonald, 2009; Wegener, 1991). Specifically, men tend to have larger networks comprised of other men and non-kin (weaker ties), while women have more kin (stronger ties) in their networks. African-Americans may have less diverse and smaller networks than whites, thereby limiting access to new information (Lin, 2000). As a result of these patterns, women and minorities tend to receive significantly fewer job leads than white men from their social networks (McDonald, 2009).

Having weak social ties is not always an indicator of the ability to obtain job information; social class and race can influence the usefulness of weak ties in job-seeking. Individuals who have held higher status jobs in the past tend to benefit from their weak ties while those who have held lower status jobs may not (McDonald, 2009; Wegener, 1991). One examination of job-seeking among the African-American urban poor in the US found that, contrary to findings regarding network size in Lin (2000), this population has large social networks but face challenges mobilizing social ties for the purposes of employment (Smith, 2005).

Access to social capital as a function of social ties can also influence job-seeking. Social capital refers to the structure and content of social relationships that provide resources including information about, or access to, job opportunities (Coleman, 1988; Portes, 1998; Lin et al., 2001). For example, African-American men are less likely to have ties with individuals in authority positions who could help them obtain higher level management positions (McDonald, 2009). Social capital is created when social connections are characterized by trust and reciprocity (Abbott, 2008; Kunitz, 2004). An individual may have social capital as a result of person-to-person ties with similar people such as family, friends, neighbors, close acquaintances (sometimes referred to as *horizontal* social capital) or through ties with individuals and organizations that are different (sometimes referred to as *vertical* or bridging social capital) (Kunitz, 2004). People who live in areas where there is less trust and reciprocity reported by residents (low horizontal social capital) and lack access to vertical social capital are likely to have fewer employment opportunities compared to those who live in high social capital areas (Hawe, 2000; Kawachi, Kennedy, & Glass, 1999; McDonald, 2009).

While there has been a great deal of attention to the role of social ties and social capital in employment, most studies focus on urban settings or have not differentiated between urban and rural residents. The focus on rural communities is particularly important in the US given high rates of poverty and unemployment in these areas, and inherent limitations on the number and types of employers (http://www.dailyyonder.com/rural-unemployment-soars-january/2010/03/30/2667). It is important to note that within the US unemployment rates vary by race and gender. The unemployment rate for black men in the US in 2010 was 18.4%, this is higher than for black women (13.8%), white men (9.6%),white women (7.7), Hispanic men (12.7%), and Hispanic women (7.7%) (Bureau of Labor Statistics, 2010).

Men on the Move (MOTM) is a community-based participatory research project located in a rural African-American community in the state of Missouri and is funded by the National Institute for Minority Health and Health Disparities. One focus of MOTM is to understand and address barriers to employment for African-American men in this community with the distal goal of improving health. As with other parts of the country, men living in rural communities in Missouri are likely to have higher rates of unemployment than their urban counterparts (http://health.mo.gov/living/families/ruralhealth/pdf/biennial2011.pdf). In order to better understand the social ties and social capital of African-American men in a rural setting, and compare their social networks to others in their community, we address three specific research questions:

1. What are the characteristics of job-seeking networks in a rural county with high unemployment?
2. How do the job-seeking networks of employed and unemployed, African-American and white, male and female residents differ?
3. What is the role of social capital in these job seeking networks?

2. Methods

Data Collection

Using the mixed-methods network approach, *net-mapping*, developed by Eva Schiffer (http://netmap.wordpress.com/), we trained four residents of rural Pemiscot County, Missouri to conduct interviews with local residents about their most recent job search. With a goal of conducting interviews with five residents from each of eight groups (Table 1), interviewers recruited 39 county residents through local churches and local unemployment and job centers.

Interviewers worked with each participant to draw a network depicting the individuals and organizations they identified in response to this prompt: *Who was involved in your most recent job search?* As the participants were answering the prompt, each individual or organization was added to a sheet of paper with the participant shown in the middle. For the purposes of describing the job-seeking networks, each participant is the ego and each individual or organization a participant identified is an alter. This approach to data collection is ego-centric; it essentially results in a network for each participant centered around the ego. Each alter was recorded on a color coded post-it note to identify whether the alter was male, female, or an organization/website. The participant then indicated whether each alter was a family member, friend, or professional tie. Participants were asked if they trusted each alter, the level of influence the alter had on their job search, and the employment status of the alter. Finally, participants were asked to indicate whether their alters knew one another. The resulting networks looked like those shown in Figure 1.

In addition to drawing their job-seeking network, each participant was asked to complete two questionnaires: one about themselves and one about their alters. Both questionnaires included demographic information. The alter questionnaire also included questions about the length of relationships in years. Strength of ego-alter relationships was also assessed by asking: *How well do you know [alter name]?* with response options ranging from 1 to 5, where 5 is "very well." To assess social capital, participants were also asked if named alters could connect them with other people or organizations in the community that might help in the job search. To obtain this information participants were asked: *Next, we are interested in the extent to which people you named have relationships with someone that might help you or others find jobs. Does [name of alter] have a relationship with the schools or parent teacher organizations?* The question was repeated for local business, community group, church, bank, school board, city council, and police department/fire department. Each mapping and survey process took approximately one hour. The Saint Louis University Institutional Review Board approved this study.

Data Analysis

Descriptive statistics and ego network visualization (e.g., Figure 1) were used to examine and compare the size, composition, and relationship structures across the job-seeking networks. To further explore patterns identified across participant networks, a novel technique using exponential random graph modeling (ERGM) to simulate a whole network from ego-centric network data was also used to examine patterns of connections across this community (Krivitsky, Handcock, & Morris, 2011). By following Krivitsky and colleagues (2011) and a 2011 tutorial by Krivitsky and others (https://statnet.csde.washington.edu/trac/raw-attachment/wiki/Resources/ERGMtutorial.pdf), aggregate data from sampled ego-networks can be used to statistically examine patterns of connections across a population. Organizations and websites were not included in this analysis since they do not have the characteristics of interest (gender, race, employment status).

To derive a whole network from the 39 ego networks, we first entered the number of network

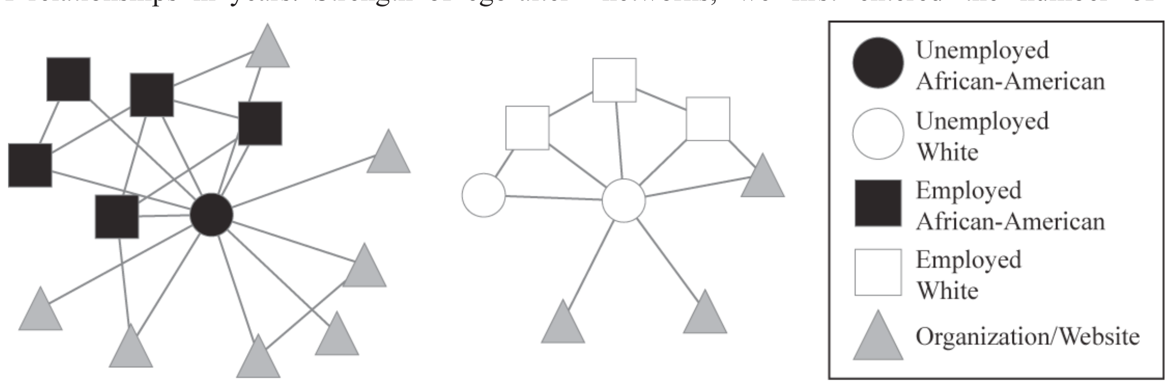


Figure 1. Examples of participant job-seeking networks showing an unemployed African-American woman (left) and an unemployed white man (right).

members for each of the eight groups (e.g., unemployed white men, employed white women, etc.) into the procedure. Because we over-sampled unemployed and African-American residents, the data were weighted so that the composition of the network would accurately represent the proportion of people in each of the eight categories among Missouri residents as of 2010 (http://www.bls.gov/cps/). State-level estimates were the closest geographic level available for estimates of race by gender by employment status. Unemployment rates in 2010 for Missouri were: black men 16.5%, black women 14.3%, white men 9.3%, and white women 8.0%. Table 1 shows the number of human egos and alters for each characteristic of interest across the 39 ego networks (n), the weighting to ensure the resulting whole network has the same composition as the state, and the resulting proportion of the whole network with each characteristic of interest (Adjusted n). The weighting column is a ratio of the proportion of the characteristic in the sample to the proportion of the characteristic among residents of Missouri.

In addition to the number of each type of network member, we entered the number of connections between each of the eight groups, also weighted using the Table 1 weights. To model the propensity for homophily (the tendency for like network members to connect with each other) by gender, race, and employment status, we entered the sum of the weighted ties that were between network members of the same employment status, race, and gender. Finally, the sum of weighted ties going to unemployed, African-American, and male network members was entered in order to model the likelihood of being sought for job advice in these three groups. In addition, because the average number of human ties for an ego was 4, we specified approximately 4 job-seeking ties per network member for the whole network.

Once the summary statistics for the number of network members and their patterns of ties were entered, a statistical model (ERGM) was generated to predict the likelihood of a tie between any two community members based on the summary statistics. We built the network starting with a null model that only accounts for the overall number of ties in the network, added homophily terms, and added variables to account for incoming ties for male, unemployed, and African-American residents.

Model fit was assessed by simulating whole networks from the null, homophily, and full models and comparing their summary statistics to the summary statistics for the observed data. Simulated networks represent job-seekers with directed ties between them. For example, for network members A and B, the relationship A -> B indicates that A contacted B. Outgoing ties, therefore, represent contacting someone and incoming ties represent being contacted. The characteristics of the simulated whole networks were examined to determine if the patterns of ties were consistent with those seen across the ego networks for each step of model building. In this case we simulated 200 networks based on each model and compared summary statistics across the simulated networks to the observed data. Once model fit was deemed acceptable, we interpreted model results and compared the results to patterns seen in the descriptive statistics for the ego networks.

3. Results

Participants were 32 years old on average and many (25.7%) had less than a high school education. Unemployed participants had lower levels of education; there was no notable difference in age between employed and unemployed participants. White women and African-American men had been unemployed the longest on average; 15 and 12.5 months, respectively.

Table 1: Weighting of network member types to represent proportion of Missouri population in the network model.

Subgroup	Observed n	Weighting	Adjusted n for model
Employed African-American men	38	.59	22
Unemployed African-American men	14	.32	4
Employed African-American women	41	.63	26
Unemployed African-American women	14	.31	4
Employed white men	29	2.14	62
Unemployed white men	12	.47	6
Employed white women	34	2.05	70
Unemployed white women	20	.27	5
Total	202		202

Describing the Networks

Network size and composition. The average number of alters in a job-seeking network was 7.4, with a range of 3-14. Networks had between 2 and 11 alters who were people with the rest of each network comprised of websites and organizations used during job-seeking. The overall size of the job-seeking networks differed by gender, race, and employment status. White men had the smallest networks, with 5.6 alters for employed white men and 6.3 alters for unemployed white men. Unemployed African-American men and unemployed African-American women had the largest networks with 9.2 alters on average. Table 2 describes the composition of an average ego-network in each group. On average, unemployed egos had a higher proportion of employed alters compared to employed egos, with the exception of employed African-American men, who had few unemployed alters.

Tie strength. Unemployed African-American women and employed African-American men had the longest and strongest relationships on average while unemployed white men had the shortest relationships and unemployed African-American men had the weakest relationships. Unemployed women had longer and stronger relationships on average than their employed counterparts, while the opposite was true for men. Alters in the job-seeking networks of men regardless of race or employment status had approximately the same number of ties to other alters, while alters in the networks of employed women had more alter-alter ties than alters in the networks of unemployed women. Having more ties among alters may be an indicator that a job seeker has a strongly connected cluster of friends, family, and others.

Alter types. Employed white women and employed white men had the most unemployed in their networks. Employed women had more family ties than unemployed women, while unemployed white men had the most family ties in their networks. The networks were relatively racially homogenous; the most race diversity was seen in the job-seeking networks of employed African-American women egos, who had 24.5% non-African-American alters. Unemployed African-American women had no white alters, while unemployed white men had only white alters. Men egos had a majority of men alters, while women egos tended to have women alters.

Social capital. Alters of employed African-American men and white women had the most community ties on average, while alters of the unemployed generally had fewer community ties. The exception to this was unemployed African-American men whose alters had more community ties than most other groups.

Statistical network modeling

A statistical model was generated based on the characteristics of the egos and alters, along with aggregate information about the patterns of ties between egos and alters across the 39 observed networks (Table 3). Homophily terms demonstrated that the likelihood of a tie increased between two community members of the same sex (OR=1.91; 95% CI: 1.59-2.29) and same race (OR=20.76; 95% CI: 15.91-27.09), but decreased between community members of the same employment status (OR=.42; 95% CI: .33-.53). In contacting others during job seeking, holding the rest of the network constant, community members were more likely to have contact with someone of the same gender and race, but less likely to have contact with someone of the same employment status. Holding all else constant, job-seekers were more than twice as likely to contact unemployed community members (OR=2.73; 95% CI: 2.08-3.59) compared to employed, and were 75% less likely to contact African-Americans (OR=.25; 95% CI: .21-.29) compared to whites. Being male did not influence the likelihood of being contacted compared to being female, all else held constant (OR=1.02; 95% CI: .86-1.21).

To assess model fit, we compared the target statistics from 200 networks simulated from each model with the observed target statistics. Averaged target statistics from the 200 simulated networks were within 6% of the observed target statistics across all the models during the model building process (Table 3). For example, the 200 simulated networks for the full model had an average of 349.1 ties for employment homophily, while the observed network data entered had 344 ties with employment homophily. Visualizing simulated networks during model building demonstrates the extensive homophily by race in this community (Figure 2). The null model simulated network (Figure 2a), which only accounted for the total number of ties in the network, shows many connections between whites and African-Americans. Simulations from the other two models demonstrate that, once homophily is accounted for, the networks appear segregated. In the full model the added terms for incoming ties appeared to increase the density of ties within race groups (Figure 2c).

When comparing the ERGM results to the descriptive statistics in Table 2, the model results capture many of the descriptive patterns demonstrated and provide additional information. For example, in Table 2, most connections were between two community members of the same race. The strong positive homophily term for race in the model captures this propensity for whites to connect to other whites and African-Americans to

Table 2: Composition of job-seeking networks for 39 residents of a rural county in Missouri

		Unemployed Men		Employed Men		Unemployed Women		Employed Women	
	Total	African-American	White	African-American	White	African-American	White	African-American	White
Network composition (n)									
Alters	7.4	9.2	6.3	6.8	5.6	9.2	7.0	7.5	7.0
People	4.4	4.4	3.5	4.0	3.4	4.0	4.8	5.7	5.0
Organizations	2.0	3.8	2.0	2.3	1.4	2.6	1.4	1.7	1.0
Websites	.9	1.0	.8	.5	.8	2.6	.8	.2	1.0
Tie strength (mean)									
Relationship length in years	18.4	17.7	8.1	20.7	22.3	22.8	12.0	17.4	11.1
Relationship strength	4.3	3.3	3.5	4.5	4.3	4.5	4.2	4.1	3.7
Alter-alter ties per alter	.8	.6	.6	.6	.6	.8	.4	1.4	1.1
Alter type (% of human alters)									
Family	52.6	52.8	79.2	54.2	41.3	42.0	42.3	69.7	50.4
Friend	84.2	91.0	91.7	70.8	85.3	100.0	82.4	88.6	64.7
Professional	16.4	1.7	0.0	25.0	17.4	0.1	6.7	26.9	32.0
Unemployed	22.8	28.6	28.8	12.5	34.7	10.0	10.5	20.7	34.9
Male	47.9	67.1	56.3	70.8	65.3	37.0	22.3	33.1	40.4
African-American	50.3	84.3	0.0	100.0	0.1	100.0	15.2	75.5	8.6
Alter social capital (mean)									
Community ties per human alter	2.9	2.4	1.3	4.2	3.8	2.5	3.0	2.5	3.7

connect with other African-Americans. In the observed data, women's networks included more alters overall and were, on average, more gender homogenous than the men's networks, resulting in more women than men being contacted during job searches. However, , after accounting for gender homophily, men were no less (or more) likely to be contacted during the job search in the network model.

4. Discussion

The job-seeking networks revealed numerous differences between unemployed and employed, African-American and white, men and women in this rural community. With the exception of white women, unemployed participants had larger job-seeking networks than employed participants. Men had more racially homogenous networks than women, and several of the groups had no race diversity in their job-seeking networks at all. Men had longer relationships overall than women while women had stronger relationships on average. The employed participants generally had higher vertical social capital (more connections to organizations in the community) than unemployed. Although the unemployed had more employed partners in their networks, vertical social capital was lower for unemployed participants. The network modeling confirmed several of these patterns, including extensive homogeneity of race in job-seeking networks. Broadly speaking, unemployed participants appeared to be well-connected, but lacking in connections to the right people and organizations to aid in their job search.

The findings suggest several strategies that may reduce employment disparities in rural communities. First, job obtainment for African-American men in particular may have less to do with the size of their job-seeking network and more to do with the specific connections these job-seekers and their personal networks have to local infrastructures. Our first recommendation is that these job-seekers purposefully seek to connect with community members who are employed by, or otherwise involved with, local organizations.

Second, based on comments made during interviews it could also be that unemployed African-American men are connected to others who are unwilling to recommend them for hire, which is consistent with past research (Smith, 2005). Our second recommendation is that local employers in rural communities provide job shadowing, apprenticeship, or other temporary opportunities so that individuals can demonstrate reliability when they lack references.

Finally, the extensive race homophily in the job-seeking networks may be a factor in reduced job opportunities for African-American men. There are few African-American owned businesses in this community (Barnidge, Baker, Motton, Fitzgerald, & Rose, 2011); most employers tend to be white. If African-Americans most often connect with other African-Americans in job-seeking, they may not have direct connections to those who have access to available jobs. While developing connections with the current business owners is important, our final recommendation is for the community to also increase opportunities for entrepreneurship by local African-American residents which could result in more access to jobs through existing patterns of connections.

Table 3: The likelihood of a job-seeking connection between two network members in a rural Missouri county (bold indicates statistically significant)

	Null OR (95% CI)	Homophily OR (95% CI)	Full OR (95% CI)
Edges	**.02 (.02-.02)**	**.01 (.01-.01)**	**.01 (.00-.01)**
Homophily			
Employment		**.20 (.17-.23)**	**.42 (.33-.53)**
Race		**8.23 (6.26-10.82)**	**20.76 (15.91-27.09)**
Sex		**1.81 (1.55-2.11)**	**1.91 (1.59-2.29)**
Incoming ties			
Unemployed			**2.73 (2.08-3.59)**
African-American			**.25 (.21-.29)**
Male			1.02 (.86-1.21)
Model fit			
% target statistics	97.7	96.4	94.1
AIC	6732.3	5957.6	5176.4

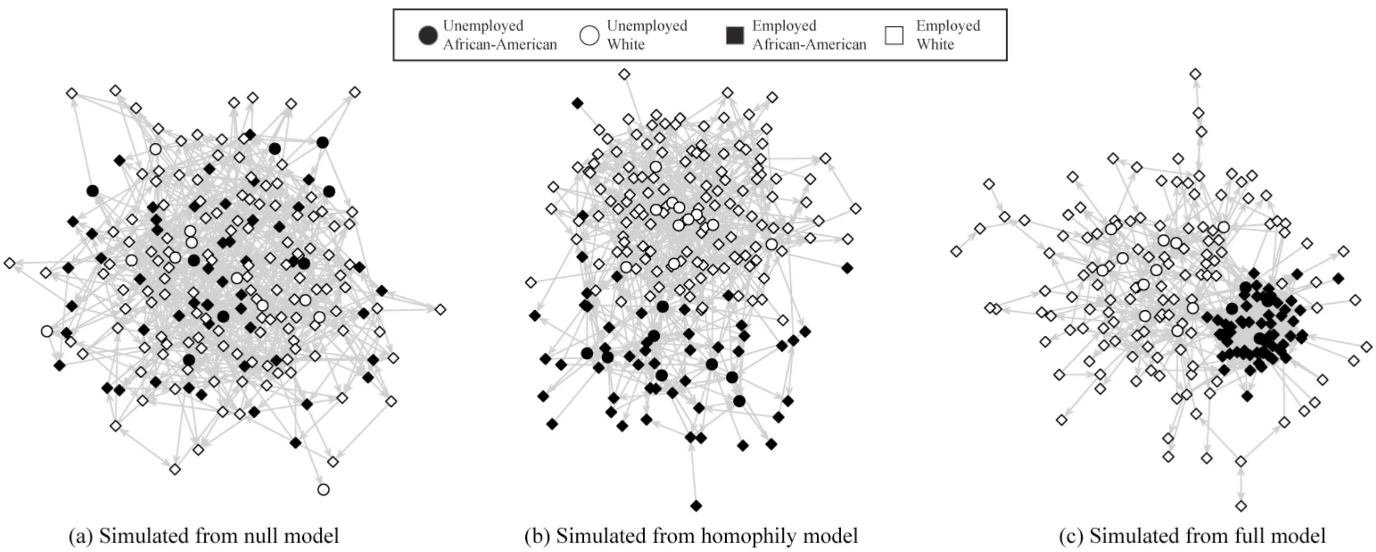


Figure 2. Simulated whole networks of job-seeking relationships among 200 residents of a rural county with high unemployment. Networks are single simulations from the (a) null, (b) homophily, and (c) full ERGM models shown *in Table 3.*

Limitations

There are several limitations that need to be considered. During recruitment, unemployed participants were selected by going to local agencies and organizations that support individuals in their job-seeking activities. While there is little information on demographic differences between those who use in-person employment agencies and those who do not, there is some evidence that, at least in urban areas, accessibility to these organizations may vary by race and age (Joassart-Marcelli & Giordano, 2006). We purposefully selected individuals based on race and average age was consistent across employed and unemployed participants, so these differences may not apply in our study. However, there may be other differences that we are not aware of between the unemployed who use services and those who do not use (or are no longer using) employment services.

Similarly, our sampling of employed individuals may be biased. We obtained our sample of employed individuals from local churches. Churches often provide strong referral networks for job seekers and may also use reputational capital to support job seeking church members (Putnam, 1993). Those who attend churches in rural communities with limited job opportunities and smaller areas of influence are likely to have stronger job referral networks compared to those who do not which may bias our sample toward the employed having more and stronger job-seeking ties than other employed individuals in the community. In addition, churches tend to lack racial diversity, especially in rural areas (Dougherty, 2003) This suggests the potential for a lack of racial diversity in the job-seeking networks of those we recruited in churches. Specifically, the employed participants recruited in churches may have more racially homogenous job-seeking networks than the general population of employed residents in Pemiscot County. The small number of egos interviewed may exacerbate these potential biases.

There are also limitations to the way we collected and analyzed our data. For example, we asked egos to identify if alters had connections to certain local infrastructures. It is possible that alters had such connections but this was unknown by the ego, or they used to have such connections but do not any longer. This may have under or overestimated vertical social capital. In terms of analysis, using state data to weight our findings likely results in underestimating unemployment which is higher in many rural areas within Missouri than the overall state rate may indicate.

Conceptually, another limitation to our approach is that we focus on network influences on employment, but do not examine the specific role of either interpersonal or structural racism (such as hiring practices or zoning) on employment opportunities within rural communities.

5. Conclusion

Increasing opportunities in this community, and other similar communities, to address high unemployment and its consequences will require effort from the job-seekers and others to develop new relationships, programs, and policies. As a first step toward implementing our recommended strategies in Pemiscot County, MOTM academic and community staff are working to aid unemployed residents in enhancing their social connections.

References

Abbott, S. (2008). Social capital and health: Starting to make sense of the role of generalized trust and reciprocity. *Journal of Health Psychology*, 13(7), 874-883.

Barnidge, E. K., Baker, E. A., Motton, F., Fitzgerald, T., & Rose, F. (2011). Exploring community health through the sustainable livelihoods framework. *Health Education & Behavior*, 38(1), 80-90.

Bureau of Labor Statistics (2010). Unemployment rates by race and ethnicity. Retrieved May 22, 2013, from http://www.bls.gov/opub/ted/2011/ted_20111005.htm.

Coleman, J. S. (1988). Social capital in the creation of human capital. *The American Journal of Sociology*, S95-S120.

Dougherty, K. D. (2003). How monochromatic is church membership? Racial-ethnic diversity in religious community. *Sociology of Religion*, 64(1), 65-85.

Friedkin, N. (1980). A test of structural features of granovetter's strength of weak ties theory. *Social Networks*, 2(4), 411-422.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1(1), 201-233.

Hawe, P. (2000). Social capital and health promotion: A review. *Social Science & Medicine (1982)*, 51(6), 871.

Jackson, P. B., & Williams, D. R. (2006). The intersection of race, gender, and SES: Health paradoxes.

Joassart-Marcelli, P., & Giordano, A. (2006). Does local access to employment services reduce unemployment? A GIS analysis of one-stop career centers. *Policy Sciences*, 39(4), 335-359.

Kawachi, I., Kennedy, B. P., & Glass, R. (1999). Social capital and self-rated health: A contextual analysis. *American Journal of Public Health*, 89(8), 1187-1193.

Krivitsky, P. N., Handcock, M. S., & Morris, M. (2011). Adjusting for network size and composition effects in exponential-family random graph models. *Statistical Methodology*, 8(4), 319-339.

Kunitz, S. S. J. (2004). Social capital and health. *British Medical Bulletin*, 69(1), 61-73.

Lin, N., Cook, K. S., & Burt, R. S. (2001). *Social capital: Theory and Research* Transaction Publishers.

Lin, N. (2000). Inequality in social capital. *Contemporary Sociology (Washington)*, 29(6), 785-795.

McDonald, S., Lin, N., Ao, D. (2009). Networks of opportunity: Gender, race, and job leads. *Social Problems (Berkeley, Calif.)*, 56(3), 385-402.

Portes, A. (1998). Social capital: Its origins and applications in modern sociology. *Annual Review of Sociology*, 24(1), 1-24.

Putnam, R. (1993). The prosperous community: Social capital and public life. *The American Prospect*, 13(Spring), Vol. 4. Available online: http://www.prospect.org/print/vol/13.

Smith, S. S. (2005). "Don't put my name on it": Social capital activation and Job-Finding assistance among the black urban poor. *The American Journal of Sociology*, 111(1), 1-57.

Wegener, B. (1991). Job mobility and social ties: Social resources, prior job, and status attainment. *American Sociological Review*, 56(1), pp. 60-71.

Dependency Centrality from Bipartite Social Networks

Luke M. Gerdes
United States Military Academy
West Point, New York

Abstract
This paper introduces dependency centrality, a node-level measure of structural leadership in bipartite networks. The measure builds on Zhou et al.'s (2007) flow-based method to transform bipartite data and captures additional information from the second mode that existing measures of centrality typically exclude. Three previously published bipartite networks serve as test cases to demonstrate the extent of correlation among node-level centrality rankings derived from dependency centrality and those derived from canonical centrality measures: degree, closeness, betweenness, and eigenvector. Ultimately, dependency centrality appears to offer a novel means to measure importance in bipartite networks depicting social interactions.

Author
Luke M. Gerdes is an Assistant Professor in the Department of Behavioral Sciences & Leadership at the United States Military Academy in West Point, New York.

Notes
This work was supported by the Office of the Secretary of Defense, Minerva Initiative. The views expressed herein are those of the author and do not purport to represent the official policy or position of the United States Military Academy, the Department of the Army, the Department of Defense, or the United States Government.

Please send all correspondence to Luke M. Gerdes, Department of Behavioral Sciences & Leadership, United States Military Academy. Email: luke.gerdes@usma.edu

1. Introduction

Bipartite data has been prominent in social network analysis since the discipline's earliest days. Indeed, Freeman (2004) traces the origins of bipartite networks to the late 1890s, when John A. Hobson, a British newspaper man and economist whose work influenced Lenin and other prominent Marxists, gave birth to the much studied topic of interlocking corporate boards. He presented agent x organization data summarizing six men's co-membership on the boards of five prominent South African companies. In pre-WWII America, Davis et al (1941) continued the tradition of bipartite data and studied social ties among southern women by cataloguing their co-participation in luncheons and parties, leading to the creation of an agent x event dataset that remains widely studied to this day.

Despite its long history, bipartite data has always presented analytic challenges. Because standard measures of node-level centrality were designed to estimate actors' influence within a single mode (Freeman, 1979), they lack direct applicability to bipartite data. There have been several efforts to extend centrality to multi-modal data (Faust, 1997; Borgatti & Everett, 1997; Latapy et al, 2008), but these approaches have yet to gain wide acceptance. The majority of analysts continue to transform bipartite data into a single mode prior to its analysis, even though it is widely known that this approach can negatively impact the theoretical validity of measurements conducted in the resulting one-mode projection (Opsahl, 2013). For example, Wasserman and Faust (1994) note that degree centrality performs differently in one-mode mode data derived from transformation than in 'naturally occurring' one-mode data; projected data tends to contain a number of abnormally large cliques resulting from the inference of direct ties among all co-participants in relatively large events.

Other scholars have offered new approaches to data transformation that partially mitigate such pernicious effects (Padrón et al, 2011; Opsahl, 2013; Gerdes, 2014). Most notably, Newman (2001) designed an approach that divides the weight of ties formed through co-participation by the number of participants in each event, which causes the resulting one-mode network to value ties formed during large events less than those formed during small events. Thus, Newman's process restricts nodes' ability to acquire high degree centrality through participation in large cliques.

Although this and other similar transformation approaches are elegant, context can undermine their utility. More manpower is often required to solve hard problems or manage complex situations than is needed to deal with small problems and everyday situations. These circumstances suggest that a clique's output may be directly proportional to its size. Consequently, large cliques are not always indicative of fleeting relationships. Even when analysts implement more sophisticated methods to transform bipartite data to a single mode, there are often still grounds to question how well measurements of centrality conducted in the resulting one-mode projection describe node-level influence and importance.

In their effort to find a more nuanced means of comparing the similarity of actors in bipartite networks, Zhou et al (2007) point the way toward a new measure of centrality that captures additional information from the second mode. This approach mitigates some of the concerns regarding the validity of node-level centrality in projected data. The new measure, which this paper dubs "dependency centrality," is based on a derivation of structural equivalence among nodes and relies on a novel method that Zhou and his co-authors devised to transform bipartite networks into single mode networks. The first section following this introduction describes Zhou's method of projection, before the second section moves-on to discuss the calculation of dependency centrality. The third section situates this new measure within the context of existing scholarship by discussing dependency centrality's relationship to other measures, especially those described in Kleinberg's (1999) work on centrality among web pages. The fourth section presents results that compare dependency centrality to existing node-level heuristics, and the fifth and final section offers some brief conclusions about measurement's potential applications.

2. Transforming Bipartite Networks

Consider a bipartite network comprised of agents and events. By considering the agents as the holders of resources that flow through the network, Zhou et al determined that it is possible to infer an agent x agent matrix from these between-mode connections. Zhou's process has two stages: in the first, the resources flow from the agents to the events in direct proportion to each agent's degree. Thus, if an agent holds links to two events, half of his resources flow to each of these events; if an agent has links to three events, one third of his resources from to each of these events, and so on.

However, the resources do not remain parked on the events. Instead, they immediately flow back to the agents following the same redistribution rules. If an event has ties to four agents, then a quarter of the resources momentarily parked on the event flow to each of the four agents who share ties to the event. If an event has ties to and so on. Mathematics allows these two processes to be combined into a single step:

$$w_{ij} = \sum_{k:\, m_{ik};\, m_{jk} = 1} 1/(\deg_j * \deg_k),$$

where w_{ij} represents that strength of the tie that exists between *i* and *j* as the result of their co-participation in *k* events in the original bipartite matrix, *M* (Gerdes, 2014). Figure 1 also offers a graphic depiction of this flow-based process for a simple bipartite network containing four individuals (Tom, Dick, Bob, and Harry) and four events (A,B,C, & D).

It is worth noting that the flow metaphor explaining this data transformation process holds regardless of the unique features of the underlying bipartite network. Figure 2, which depicts a simple network containing an exceptionally skewed distribution of ties, offers a useful illustration. In the one-mode projection that results from the application of Zhou's process to this skewed network, every actor in the network receives an equal inbound tie to Tom, Dick, Harry, and Bob. However, because Ron participated in every event in the network, and because he was the only participant in events B through C, a smaller share of his resources flow to Tom, Dick, Harry, and Bob, while the majority of Ron's resources return to him. Thus, the redistributions inherent to Zhou's flow-based approach are not necessarily egalitarian. Instead they operate according to each network's unique underlying structure, which causes the redistributions in the network depicted in figure 2 to function akin to an exceptionally regressive tax: the resources held by the poorest agents (i.e. those with low degree in the two-mode network) are taken and redistributed evenly. The resources held by the single rich agent (i.e. the lone high-degree agent in the two-mode network) largely return to him, even as he receives an equal share of the resources the poorer agents held in the original bipartite network.

While the logic of Zhou's approach holds regardless of the specifics of the underlying two-mode network, astute observers will notice that this transformation algorithm only allows for binary data. Fortunately, this limitation can be overcome using the "pipes" approach that Newman (2004) utilized in his efforts to calculate second generation measures of centrality. The basic logic is simple: when nodes share an interaction of weight greater than 1, analysts should think of each additional unit of weight as an extra pipe between the nodes. If all pipes in the analysis are of equal diameter and are equally full, then the value of ties between ego and any alter can be expressed as a ratio of the total number of pipes that ego 'owns.' Thus, in figure 3, Tom sends a tie of 5/6 to Event A, because five pipes connect these two nodes, and Tom owns a total of 6 pipes. Similarly, in the second step of the transformation, Event A sends 4/9 of its resources to Harry, because four pipes connect these two nodes, and Event A 'owns' a total of 9 pipes. Using this convention, a generalization of Zhou's transformation process that allows for weighted data can be formalized as:

$$w_{ij} = \sum_{k:\, m_{ik};\, m_{jk} > 0} (m_{ik} * m_{jk})/(\deg_j * \deg_k),$$

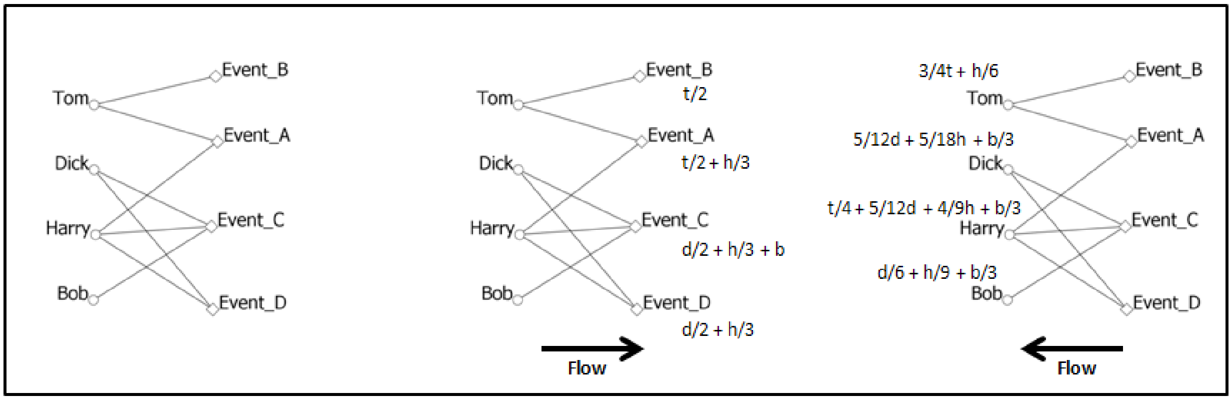
where w_{ij} is the strength of the relationship that agents *i* and *j* share as the result of co-participation in *k* events in the original two-mode matrix, *M* (Gerdes, 2014). Figure 3 offers a graphic depiction of the "pipes" generalization of Zhou's transformation process.

3. Toward Dependency Centralization

Even when generalized to accommodate weighted data, Zhou's process has limitations for analysts wishing to study traditional measures of centrality. Specifically, the method flattens the distribution of degree centrality by discounting the weight of new ties formed by high-degree agents. When applied to longitudinal data, the process would, consequently, undervalue the importance of novel partnerships, especially those forged to actors with high degree centrality. Moreover, the process produces directed graphs, which may not be appropriate in several circumstances involving agent *x* event data. For example, it is difficult to see how Tom could have interacted with Harry fewer times than Harry interacted with Tom through participation in the same event. It is worth noting that this criticism of directionality may become moot if the context of analysis dictates that agents hold different quantities of resources in the original bipartite graph, thereby enabling an exchange of "50 units" to be expressed as one-half of the resources held by an agent who started the transformation in possession of 100 units, and two-thirds of the resources held by an agent who started the transformation process in possession of 75 units.

Agent x Event

	Event A	Event B	Event C	Event D
Tom	1	1	0	0
Dick	0	0	1	1
Harry	1	0	1	1
Bob	0	0	1	0



Agent x Agent

	Tom	Dick	Harry	Bob
Tom	3/4	0	1/6	0
Dick	0	5/12	5/18	1/3
Harry	1/4	5/12	4/9	1/3
Bob	0	1/6	1/9	1/3

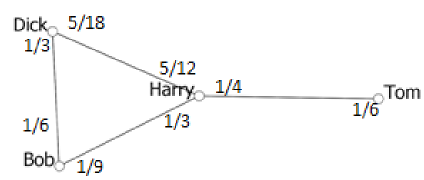
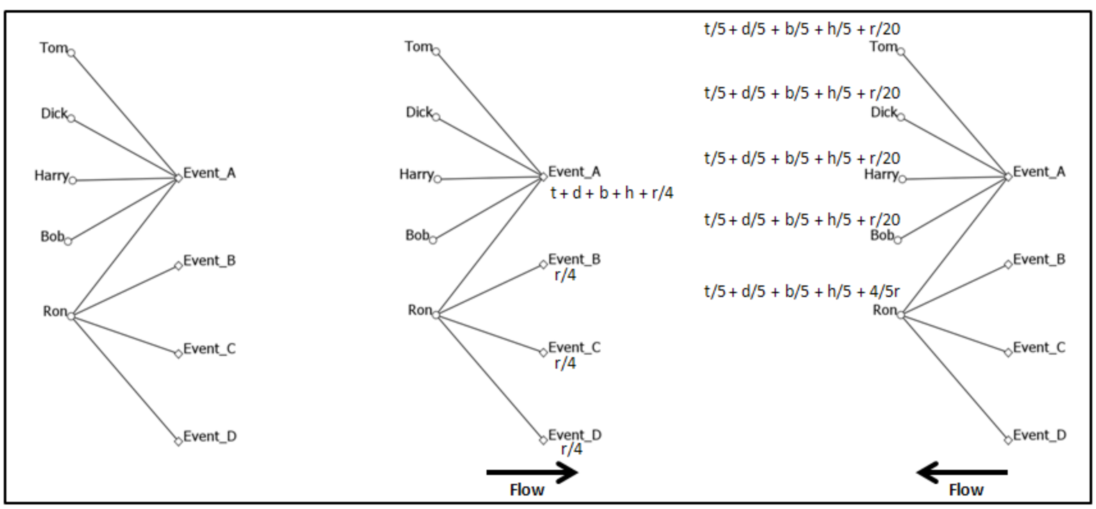


Figure 1: Zhou et al's Projection Process for Bipartite Data (Gerdes, 2014)

Agent x Event

	Event_A	Event_B	Event_C	Event_D
Tom	1	0	0	0
Dick	1	0	0	0
Harry	1	0	0	0
Bob	1	0	0	0
Ron	1	1	1	1



Agent x Agent

	Tom	Dick	Harry	Bob	Ron
Tom	1/5	1/5	1/5	1/5	1/20
Dick	1/5	1/5	1/5	1/5	1/20
Harry	1/5	1/5	1/5	1/5	1/20
Bob	1/5	1/5	1/5	1/5	1/20
Ron	1/5	1/5	1/5	1/5	4/5

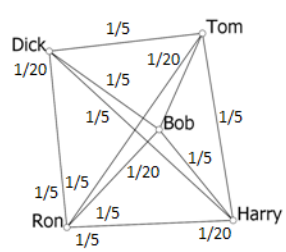
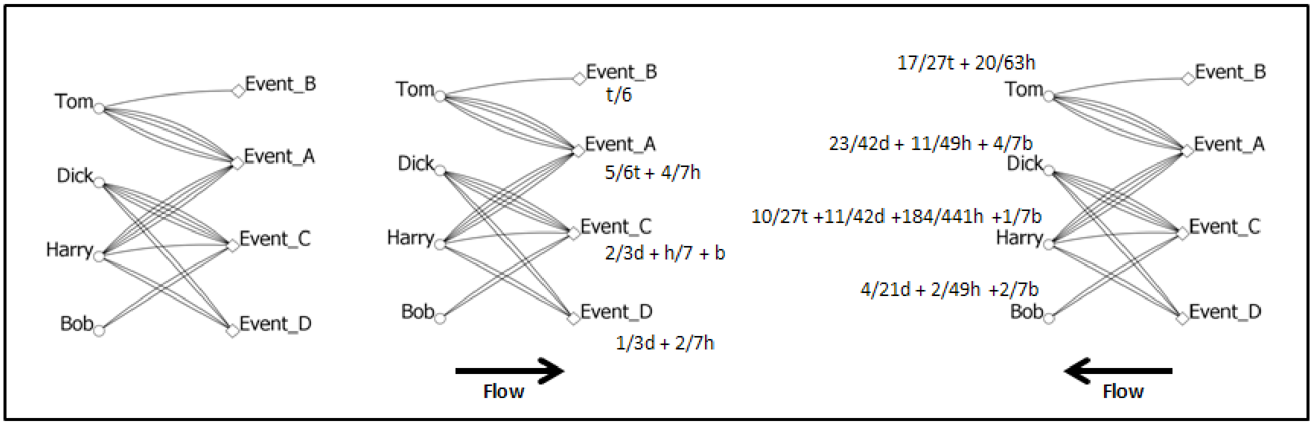


Figure 2: Zhou et al's Projection for Irregular Bipartite Data

Agent x Event

	Event A	Event B	Event C	Event D
Tom	5	1	0	0
Dick	0	0	4	2
Harry	4	0	1	2
Bob	0	0	2	0



Agent x Agent

	Tom	Dick	Harry	Bob
Tom	17/27	0	20/63	0
Dick	0	23/42	11/49	4/7
Harry	10/27	11/42	184/441	1/7
Bob	0	4/21	2/49	2/7

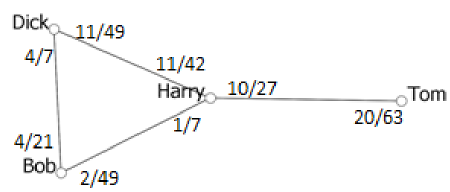


Figure 3: "Pipes" Generalization of Zhou et al's Projection Process for Bipartite Data (Gerdes, 2014).

Despite these limitations, Zhou's approach has one advantage unavailable from other methods of data transformation: this projection process enables analysts to calculate agent's structural dependence on one another. Structural dependence can be expressed as a ratio, varying between 0 (complete independence) and 1 (complete dependence), and which can be determined by dividing the number of ties an agent shares with an alter by the original agent's self-loop (i.e. the amount of resources that return to the agent at the end of the flow-based transformation). More formally:

$$D_{ij} = w_{ji}/w_{ii} ,$$

where D_{ij} represents agent *i*'s dependence on agent *j* as the result of co-participation in events.[1]

Although Zhou et al's discussion of dependency ratios does not reference structural equivalence as introduced by Lorrain and White (1971), the two concepts are highly related. Nodes *i* and *j* are structurally equivalent if they possess identical in-ties and identical out-ties. In an agent *x* event network, node *i* is structurally dependent on node *j* if the later participates in *at least* the same events as the former. If *j* participates in additional events that do not involve *i*, the dependency of *i* on *j* remains unaffected, though *j* would then be only partially dependent on *i*. Thus, dependency ratios are effectively directed, non-reciprocal measurements of structural equivalence that can only be operationalized in bipartite networks that have been transformed using Zhou et al's flow-based transformation process to infer the uniquely weighted self-loops on which dependency ratios rely.

How are these ratios useful to analysts? According to Zhou et al, the dynamics of the professor-student relationship offer a useful illustration. If a professor co-authors a series of papers with several different graduate students, none of whom have previously published, each student will be entirely dependent on the professor (i.e. each student's *D* will reach the theoretical maximum of one), because they have no other publication partners.

[1]Zhou et al originally named this measure "independence," but as higher values indicate greater levels of dependence, their naming convention was counter-intuitive. The calculation remains unchanged from the original formulation, but has been re-titled to remain in-line with typical network conventions, as best demonstrated by the fact that greater indirect influence is typically conceptualized as an a higher measure of closeness centrality, rather than a lower measure of average path-length.

However, the professor will be independent of any single student (i.e. the professor's *D* will approach the theoretical minimum of zero) by virtue of his myriad publication partners. Thus, dependency ratios begin to capture notions of relative power.

Consequently, even in their raw state, these calculations are useful for assessing patterns across a host of applications. The student-mentor example highlights the applicability of dependency ratios to the study of citation networks, but the calculations can also help to improve the recommendations that firms provide to consumers based on the purchasing behavior of other similar customers (Zhou et al, 2007; Liu et al, 2009; Li et al, 2009). This approach may also offer a means to detect community structure in complex networks (Pan et al, 2010). Finally, as this paper contends, dependency calculations may serve to highlight leadership patterns within organizations, by assessing the extent to which supposed subordinates are structurally dependent on an organization's named leader.

Table 1: Pair-wise Dependencies

	T	D	H	B
T	0	0	0.333	0
D	0	0	1	0.400
H	0.375	0.625	0	0.25
B	0	1	1	0

In order to highlight these sorts of organizational dynamics, it is necessary to move from raw dependency calculations to a normalized node-level measurement of centrality. This task can be accomplished by conducting the sort of pair-wise comparison of dependency calculations listed in table 1, which shows all possible dependencies among participants in the one-mode network that appears in the lower left-hand corner of figure 1. This comparison ignores self-loops, since every actor is perfectly dependent on themselves.

Zhou et al propose that a node-level measurement of dependence can be determined by squaring the value of each cell, and then summing by column for each actor, such that overall dependence can be formalized as:

$$\sum_{j} (w_{ji}/w_{ii})^2 \, .$$

However, this non-linear measure is somewhat problematic. Given that it is impossible to generate a negative dependency ratio, it remains unclear why it is appropriate to square each measurement. Since cell values range between 0 and 1, this action seems to artificially inflate differences between actors. Values at the theoretical minimum (i.e. zero) and the theoretical maximum (i.e. one) remain unaffected by the squaring process, while decimalized values between these extremes become smaller when multiplied against themselves. For example, Tom's raw comparison to Harry, initially valued at one-third, becomes one-ninth when squared, while Dick and Bob's comparison to Harry, which are both valued at one, retain their original value despite squaring. Simply put, squaring can distort results for partially-dependent actors, by making such individuals appear less bound to structurally independent 'mentors' than raw dependency calculations would suggest.

The node-level measurement proposed by Zhou et al is also problematic because it lacks comparability across networks. This deficiency can be illustrated by amending the student-mentor example used to demonstrate the interpretation of dependency calculations. Assume that the goal of a study is to measure the annual change in a professor's dependence on students. Unless the professor under analysis publishes with the same number of students each year, his maximum possible dependency score will fluctuate based on the number of yearly co-authors. If he authors with five students in the first year of the study, the professor's maximum node-level score will be five, but if he authors with four students in the study's second year, his maximum node-level score will drop to four. Therefore, if the professor held a composite score of 0.5 in both years, it would be incorrect to conclude that his dependence on students was stable from year-to-year, because this score represents 2.5 percent more of the theoretical maximum in the second year ((0.5/5) – (0.5/4) = 0.025). The node-level measurement proposed by Zhou et al is ultimately of limited utility because it fails to take the size of the network into account.

Fortunately, a more intuitive, normalized node-level measurement can be calculated from the sort of pair-wise comparisons of dependency ratios listed in table 1. Begin by taking the column sums of the pair-wise dependency table. This process yields a raw cumulative measurement of the extent to which all other nodes are dependent on a given actor. For example, Harry holds a raw cumulative value of 2.33, because Tom is 1/3 dependent on Harry, while Dick and Bob are both entirely dependent on Harry. Next, divide this column sum by *n-1* in order to determine the mean extent to which all other nodes in the one-mode network are dependent on a given actor. Harry's score now becomes approximately 0.78. Freeman (1979) implemented the standard node-level normalization process division by *n-1* to ensure that centrality scores derived from binary data range between the added benefit of cross-comparability among networks of different sizes. This “dependency centrality,” which is effectively an implementation of in-degree centrality on matrix that has been processed via Zhou’s flow-based transformation algorithm, is computed by:

$$D_C = \frac{\sum_{j \neq i} \frac{w_{ji}}{w_{ii}}}{(n-1)},$$

where D_c represents dependency centrality; w_{ji} represents the weight of a tie from actor *j* to actor *i*; w_{ii} represents the weight of actor *i*’s self-loop, and *n* represents the total number of agents in the network.

When viewed in the context of bipartite agent *x* event data, this measurement seems to capture structural leadership, since agents who score highly in dependency centrality have a large number of followers who participate in events only alongside the central individual. Conversely, individuals with low, but non-zero, dependency centrality scores appear to be “followers,” who typically participate in events alongside other more active individuals. Agents who are non-isolates in the original two-mode network, but who display dependency scores of zero appear to be lone wolves, who operate absent both social leaders and social followers.

These interpretations suggest that dependency centrality offers a useful means of identifying key actors from bipartite data, which is among the most common classes of network data. However, as the following section on the relationship between this measure and existing scholarship makes clear, dependency centrality is best suited for bipartite networks that depict some aspect of social behavior. Other measures appear better suited to study bipartite networks that exist outside the human environment.

4. Relationship to Related Work

In addition to dependency centrality’s aforementioned relationship to structural equivalence, the new measure brings to mind Kleinberg’s (1999) study of the role network structure plays in the search for authoritative web pages—a work that dismisses in-degree as a valid means for finding authoritative nodes. Specifically, Kleinberg argues that in-degree can inflate the importance of superfluous web pages because many hyperlinks, such as those that allow users to navigate back to an organization’s main page, are not topically-oriented and thus have nothing to do with the interrelationships among pages containing similar content. Moreover, when “authority” is defined among a sub-set of pages that contain a specific user-determined search-term, bad tagging can further problematize in-degree’s ability to capture the concept, because the top returns of searches ordered by in-degree often fail to match the user’s intent. Kleinberg illustrates this point by noting that Amazon.com featured among pages with the largest number of in-links for pages that contained ‘information’ on the “Java” computer language. This result occurred because the breadth of Amazon’s retail catalogue makes the site “universally popular,” causing it to “have large in-degrees regardless of the underlying query topic” (p. 610-611). Such considerations led Kleinberg to conclude that eigenvector-based heuristics, which attribute importance to nodes based on the structural importance of their alters and not merely these alters’ number, offer better means to locate authoritative web pages than the naïve results obtains from in-degree.

Although Kleinberg is concerned with one-mode networks of web pages, as opposed to bipartite networks, there is a temptation to assume that his conclusions about the relative merits of in-degree and eigenvector centralities are generalizable to all networks, regardless of their underlying content and structure. But context matters. Bipartite social networks appear to be largely devoid of the characteristics that led Kleinberg to dismiss in-degree as a valid measure of node-level importance.

Consider an agent *x* event network depicting social functions. Provided that the analyst has given some thought to issues of data collection and is not attempting to infer centrality based on co-attendance of the Super Bowl, stadium concerts, or other events that boast large crowds, it is difficult to think of an example in which a person’s role in the social network is divorced from the extent of their participation in events along the same lines that caused Amazon.com to appear falsely central when Kleinberg used in-degree to rank web pages associated with the search-term “Java.” People have limited capacity to attend social events, and thus do not typically form false associations with events in the same manner or on the same scale as monolithic websites that boast virtually limitless capacity for tie formation, allowing them to hold indiscriminate associations with topics of every variety. If analysts follow sound data collection protocols in building a bipartite social network, then they can have confidence that any agent who participates in every event in the network plays an important social role.

Similarly, properly designed data collection protocols will bypass information that is akin to the navigational shortcuts (e.g. “return to home” hyperlinks) that problematize efforts to use in-degree to rank web pages based on the structure of their ties to other online portals. Returning to the example of an agent x event network depicting social functions, an analyst would only need to worry about the social equivalent of non-thematic navigational ties if the data neglected to distinguish between social functions and other types of gatherings, such as corporate meetings. The previous examples make a simple but important point: different analytic tools are appropriate for different analytic contexts. Just because in-degree functions poorly for online networks does not mean that this type of centrality functions poorly for social networks.

It is also worth examining the other side of the coin by briefly discussing the appropriateness of implementing on dependency matrices the sort of eigenvector-based heuristics that Kleinberg recommends to rank the results of online search queries. Although some prominent network theorists have recommended against the use of eigenvector-based centralities on directed data (Bonacich 1972, Valente et al, 2008), such measurements can be conducted on any square matrix, and a number of approaches have been developed to examine in-links and out-links (see Bjelland et al, 2010, for a review). Thus, mathematics does not preclude the calculation of various eigenvector-based centralities on the inherently non-symmetric dependency matrices.

However, directed versions of eigenvector typically involve transforming the adjacency matrix under study into a stochastic matrix, which transposes the position of the rows and columns, before transforming cell values through division by individual column totals (Page et al, 1998). Given that the dependency matrix is already a derivative of the derivative of the original bipartite data, it is not entirely clear what sort of information an eigenvector-based measurement of centrality would capture when applied to the dependency matrix. The resulting measurements may simply be too far removed from the original data to accurately characterize agents' influence and importance.

A more intuitive result could be obtained by transforming the original bipartite matrix using a process other than Zhou's, and then implementing an eigenvector-based measure of centrality in that matrix. The following section on testing implements this approach to eigenvector-based heuristics in its effort to compare the results of dependency centrality to several other existing measures of node-level importance.

5. Evaluating Dependency Centrality

The nature of dependency centrality complicates comparative evaluations. Because the new measure relies on the positive self-loops derived from the projection process innovated by Zhou et al, this measurement explicitly considers information that originates in a second mode. Therefore, it is intended to be applied only to bipartite networks. Consequently, dependency centrality lacks the near-universal generalizability of traditional measures of centrality (e.g. degree, closeness, and betweenness centralities), which were originally intended to measure node-level influence within a single mode (Freeman, 1979). Simply put, dependency centrality is conceptually distinct from the canonical measures of centrality, and the new measure is difficult to compare to these centralities because they attempt to measure different things.

This theoretical distinction has practical consequences for evaluation. As discussed in a preceding section, it is difficult to have full confidence in any standard measures of centrality derived from the one-mode networks resulting from the flow-based projection process of Zhou et al. Efforts to gauge the relative performance of dependency centralization must, therefore, rely on two distinct transformation processes: one to calculate standard network measures, and another that is only useful for calculating dependency centrality. Stated differently, it is necessary to derive benchmarks for comparison by processing a bipartite network using a conventional approach to data transformation, before deriving dependency centrality scores by re-processing the same bipartite network using the flow-based projection process of Zhou et al.

This paper implements a two-pronged approach to testing in regards to three previously published data sets. These are drawn from various subject matter in an effort to situate dependency centrality in a broad intellectual context accessible to researchers from diverse disciplines. First, the paper evaluates the data that Davis et al (1941) collected on the participation of 18 southern women in 14 social events (see also, Breiger, 1974). Figure 4 offers a visualization of this network (generated via Csardi & Nepusz, 2006). Next, the paper evaluates the performance of dependency centrality within the context of dark networks by testing the measure using agent x event data that Center for Computational Analysis of Social and Organizational Systems, a research group at Carnegie Mellon University, collected on the participation of 18 al-Qaida members in 25 functional tasks underlying the 1998 bombings of the U.S. Embassies in Nairobi, Kenya, and Dar es Salaam, Tanzania (CASOS, 2008; Gerdes, 2008). Figure 5 offers a visualization of this network (generated via Csardi & Nepusz, 2006). Since the first two sets of test data are binary and relatively small, the final evaluation tests the performance of dependency centrality within the context of a larger, weighted bipartite network. Data that Opsahl (2013) collected on "Facebook-like forums" utilized by students at University of California at Irvine serves this purpose. In this dataset, 899 students are tied to 522 topical forums, and tie weights are determined by the number of times that users posted to each forum. Figure 6 offers a visualization of this network (generated via Csardi & Nepusz, 2006). In all three visualizations, white circles represent agents, and gray squares represent agents.

With the selection of test cases complete, data processing was accomplished using original scripts written in the R Language and Environment for Statistical Computing (R Core Team, 2012). In addition to the flow-based transformation process that was applied to all three of the bipartite networks used in testing, the southern women and al-Qaida datasets were 'folded' into symmetrical agent x agent networks, by multiplying the original bipartite networks against their transpose. Since the application of matrix multiplication to weighted bipartite networks produces one-mode data containing badly distorted tie-strengths (Padrón et al, 2011), a "one-way sum" projection was applied to the data on "Facebook-like forums" in order to produce a weighted agent x agent network for testing.

As its name suggests, the one-way sums approach to projection works via addition. The strength of agent x agent ties is determined by conducting a pair-wise comparison of agents' weighted two-mode interactions. Specifically, this projection process assumes that an agent "sends" ties to all of the other agents who co-participate in the same event as the sender; if agent A held a tie of strength 1 to an event, and agent B held a tie of strength 3 to the same event, then the tie from A to B would be valued at 1, and the tie from B to A would be valued at 3. When individuals co-participate in multiple events, the total weight of the agent x agent relationship is found by summing across events. Thus, the one-way sums projection that was applied to the forums data can be formalized as:

$$w_{ij} = \sum_{k:\, m_{jk} > 0} m_{ik} \,,$$

where w_{ij} is the weight between node *i* and node *j* in the resulting agent *x* agent matrix, and *k* represents the event(s) in which *i* and *j* co-participated in the two-mode matrix, *M* (Opsahl, 2009; Padrón et al, 2011; Gerdes, 2014).

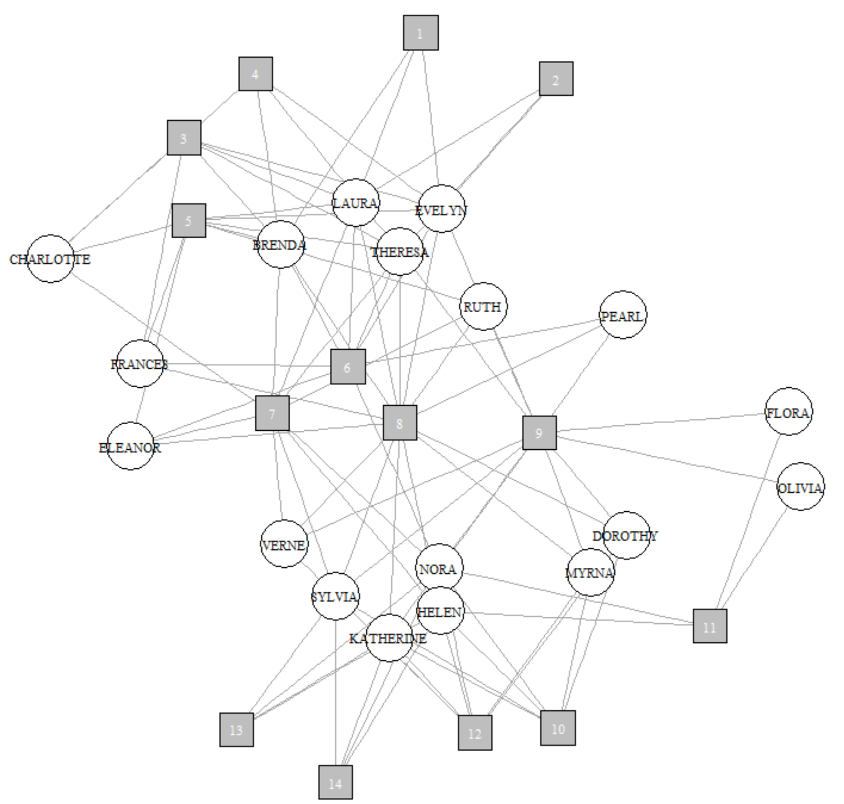


Figure 4: 18 Southern Women's Participation in 14 Events

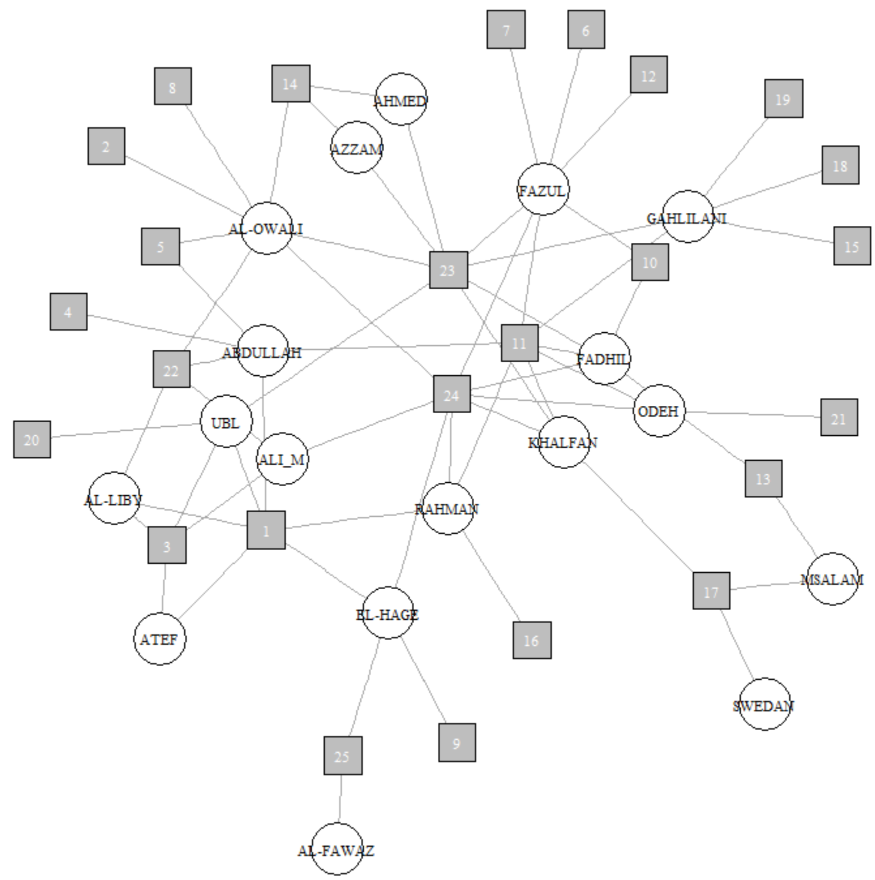


Figure 5: 18 al-Qaida Members' Participation in 25 Events

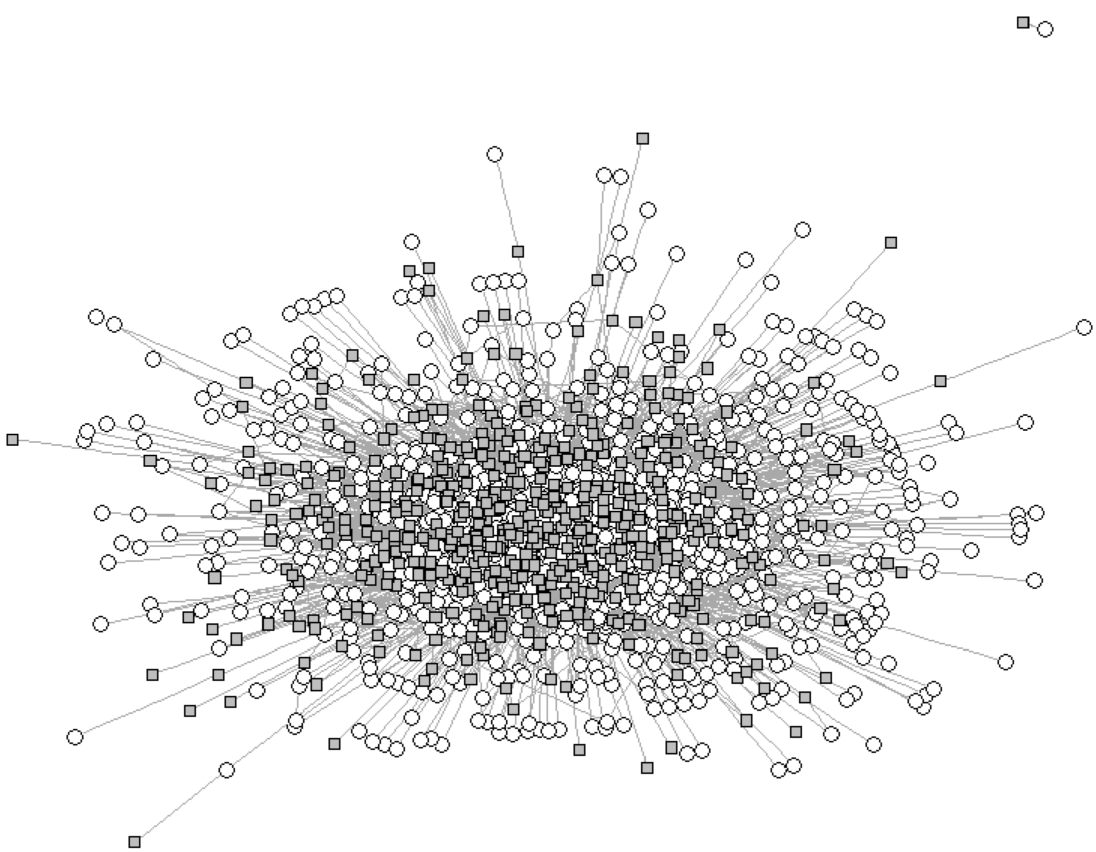


Figure 6: 899 Students' Participation in 522 Facebook-like Forums

Beyond data processing, this paper also utilized the R language to measure node-level centrality. Csardi and Nepusz's igraph package (2006) provided normalized measurements of degree, closeness, and betweenness in the one-mode networks derived from folding and one-way sums. Igraph also provided eigenvector-based measures for the networks derived via conventional approaches to transformation. Bonacich's eigenvector centrality, which was designed for symmetric matrixes, was applied to the networks derived from folding. However, because the one-way sums process is inherently directional, a fairer comparison of dependency centrality's relative performance could be obtained by applying an inherently directional eigenvector-based heuristic to the one-way sums transformation of the "Facebook-like" data. Page et al's (1999) algorithm, which underlies Google's approach to ranking the returns of online searches based on web pages' inbound links, serves that purpose. Finally, an original script was applied to the one-mode networks derived from Zhou et al's flow-based transformation process in order to measure dependency centrality.

Tables 2, 3, and 4 respectively summarize results for the southern women data, the al-Qaida data, and the forums data. These tables present both the raw scores and node rankings, and results are ordered by nodes' rank in dependency centrality. Given the size of the forums data, table 4 presents information on only the 20 individuals who rank highest in dependency centrality.

The results presented in tables 2 through 4 demonstrate that dependency centrality identifies different top-ranking agents than the canonical measurements of centrality. However, these differences tend to flatten-out for low-ranking individuals, especially when comparing dependency centrality with degree centrality or eigenvector-based centralities. Indeed, these three measures identically ranked the bottom 5 of the 18 southern women studied by Davis et al. Thus, dependency centrality performs differently than existing measures in the selection of "key" individuals and provides a unique measurement of node-level importance that is not captured by existing centrality measures.

However, it is also useful to consider the larger picture and assess the overall extent of correlation between the new measure and existing conceptions of centrality. A whole-network measure of centrality correlation is best accomplished by comparing agents' rank, since it is unsurprising that different measurements produce different raw values. Therefore, this paper relied on Spearman's Rho, which is akin to a Pearson correlation for ranks (McDonald, 2009a). Table 5 summarizes the extent of rank-correlation among dependency, degree, closeness, betweenness, and eigenvector centralities across all three of the test networks. Critical values of Rho were determined by following the widely accepted practice of approximating true values by using Student's t distribution with $df = N - 2$ (Ramsey, 1989).

Table 2: Comparing Measures of Centrality for Southern Women

Node ID	Depend.	Rank (Dep.)	Degree	Rank (Deg.)	Close.	Rank (Cls.)	Between.	Rank (Btw.)	Eigen.	Rank (Eigen.)
NORA	0.6026	1	5.5294	4	1.0000	4	0.0239	3	0.8099	6
THERESA	0.5763	2	6.7059	1	1.0000	4	0.0421	1	1.0000	1
EVELYN	0.5395	3	5.8824	2.5	1.0000	4	0.0360	2	0.8961	2
SYLVIA	0.5027	4	5.8824	2.5	1.0000	4	0.0150	5	0.8717	3
HELEN	0.4999	5	4.9412	7	1.0000	4	0.0138	6	0.7478	7
BRENDA	0.4637	6	5.4118	5	0.8947	13.5	0.0062	9	0.8518	4
LAURA	0.4510	7	5.2941	6	0.8947	13.5	0.0047	10	0.8415	5
KATHERINE	0.4203	8	4.8235	8	0.9444	9.5	0.0046	12	0.7268	8
VERNE	0.3224	9	4.5882	10	1.0000	4	0.0134	7	0.6920	10
RUTH	0.3218	10	4.7059	9	1.0000	4	0.0217	4	0.7138	9
MYRNA	0.3167	11.5	4.1176	12.5	0.9444	9.5	0.0046	12	0.6189	12.5
DOROTHY	0.3167	11.5	4.1176	12.5	0.9444	9.5	0.0046	12	0.6189	12.5
ELEANOR	0.2951	13	4.2353	11	0.8947	13.5	0.0031	14.5	0.6754	11
FRANCES	0.2787	14	3.7647	14	0.8947	13.5	0.0031	14.5	0.6087	14
PEARL	0.2402	15	3.6471	15	0.9444	9.5	0.0092	8	0.5595	15
CHARLOTTE	0.2329	16	2.8235	16	0.7391	18	0.0000	17	0.4810	16
OLIVIA	0.1659	17.5	1.6471	17.5	0.7727	16.5	0.0000	17	0.2322	17.5
FLORA	0.1659	17.5	1.6471	17.5	0.7727	16.5	0.0000	17	0.2322	17.5

Table 3: Comparing Measures of Centrality for al-Qaida Members (18 Agents)

Node ID	Depend.	Rank (Dep.)	Degree	Rank (Deg.)	Close.	Rank (Cls.)	Between.	Rank (Btw.)	Eigen.	Rank (Eigen.)
AL-OWALI	0.2531	1	2.3529	4	0.8095	3	0.0982	5	0.8473	4
KHALFAN	0.1943	2	2.5882	1.5	0.8500	1	0.3576	1	0.9513	3
UBL	0.1900	3	1.7647	6.5	0.8095	3	0.1171	4	0.5711	9
FADHIL	0.1701	4	2.5882	1.5	0.8095	3	0.1428	3	1.0000	1
FAZUL	0.1590	5	2.4706	3	0.7727	6.5	0.0388	8	0.9937	2
EL-HAGE	0.1572	6	1.5294	9.5	0.7727	6.5	0.2599	2	0.5334	11
ABDULLAH	0.1505	7	1.7647	6.5	0.7727	6.5	0.0468	7	0.6420	7
AL-LIBY	0.1504	8	1.2941	12	0.5862	14	0.0029	13	0.3923	14
ALI_M	0.1350	9	1.5294	9.5	0.7391	9	0.0346	9	0.5571	10
RAHMAN	0.1290	10	2.1176	5	0.7727	6.5	0.0599	6	0.8034	5
MSALAM	0.1069	11	0.3529	16	0.5000	16	0.0037	11.5	0.1260	16
ATEF	0.0933	12	0.9412	15	0.5667	15	0.0000	16	0.2787	15
GHAILILANI	0.0927	13	1.5294	9.5	0.6800	10.5	0.0106	10	0.6406	8
AHMED	0.0912	14.5	1.0588	13.5	0.6071	12.5	0.0000	16	0.4175	12.5
AZZAM	0.0912	14.5	1.0588	13.5	0.6071	12.5	0.0000	16	0.4175	12.5
ODEH	0.0728	16	1.5294	9.5	0.6800	10.5	0.0037	11.5	0.6692	6
SWEDAN	0.0505	17	0.2353	17	0.4857	17	0.0000	16	0.0673	17
AL-FAWAZ	0.0164	18	0.1176	18	0.4474	18	0.0000	16	0.0333	18

Table 4: Comparing Measures of Centrality for Users of a Facebook-like Forum (20 of 899 Agents)

Node ID	Depend.	Rank (Dep.)	Degree	Rank (Deg.)	Close.	Rank (Cls.)	Between.	Rank (Btw.)	Page	Rank (Page)
100	2.4341	1	91.3619	1	0.3037	1	0.1751	1	0.0098	1
18	2.2034	2	39.6080	2	0.2912	9	0.0535	4	0.0050	9
275	1.0471	3	29.4410	5	0.2837	29.5	0.0187	14	0.0032	44
67	0.9920	4	31.4666	4	0.3007	2	0.0948	2	0.0075	2
102	0.8776	5	17.7227	17	0.2719	148	0.0068	56	0.0019	159
47	0.8764	6	25.8797	6	0.2903	10	0.0324	7	0.0053	8
626	0.8446	7	22.2082	11	0.2900	11	0.0292	9	0.0044	13
17	0.8273	8	23.7528	7	0.2803	53	0.0150	20	0.0033	38
325	0.8110	9	19.4076	14	0.2773	77.5	0.0131	24	0.0027	64
290	0.8059	10	23.6125	8	0.3002	3	0.0805	3	0.0064	4
810	0.7255	11	35.1726	3	0.2942	7	0.0452	5	0.0056	6
287	0.6853	12	22.6604	10	0.2821	38.5	0.0107	30	0.0036	27
358	0.6523	13	23.5412	9	0.2807	47.5	0.0119	27	0.0034	36
173	0.5837	14	13.3641	31	0.2741	116.5	0.0109	28	0.0025	80
195	0.5680	15	21.6292	12	0.2823	37	0.0172	15	0.0032	45
172	0.5669	16	14.6314	24	0.2845	27	0.0190	13	0.0038	21
582	0.5633	17	11.1102	39	0.2795	57	0.0206	11	0.0031	48
625	0.5179	18	17.1481	19	0.2803	53	0.0171	16	0.0031	49
206	0.5046	19	17.7684	16	0.2866	19	0.0202	12	0.0046	11
16	0.4717	20	12.5690	35	0.2785	65	0.0035	108	0.0025	75

Bonferroni's correction for multiple comparisons was also applied (McDonald, 2009b) in order to determine that each of the upper three 4 x 4 tables listed below is significant at an alpha of 0.05.

The results of the correlation analysis largely confirm what previous studies have already shown: measures of centrality are often highly correlated (Coleman et al 1966; Burt, 1987; Bolland, 1988; Faust, 1997; Valente & Forman, 1998; Lee, 2006; Valente et al, 2008). Dependency centrality is, by mean across the three test networks, most closely correlated with degree centrality, at 90.5 percent. Betweenness performs similarly and is 87.9 percent correlated with dependency centrality, while closeness is only 83.8 percent correlated with dependency centrality. When Page's approach is treated as comparable with Bonacich's approach, eigenvector-based measures of centrality are only 84.5 percent correlated with dependency centrality.

It is worth noting that these patterns did not hold across the three individual test cases that contributed to this average. In the al-Qaida data, degree was less correlated with dependency centrality than either closeness or betweenness, and eigenvector centrality performed aberrantly, scoring only a 67.1 percent correlation with dependency centrality, even though comparable correlations were larger than 90 percent in the other two test networks. Although these results dictate that it is difficult to determine exactly how closely dependency centrality mimics the results of other measures of centrality, it is clear that the new measure performs differently. This finding holds both in terms of the selection of key actors as well as for correlations of node-level importance that consider all agents in the network. Ultimately, dependency centrality provides a measurement of structural leadership that is not fully captured by existing measures.

Table 5: Correlations of Node-Level Rank as Determined by Dependency, Degree, In-Degree Closeness, Betweenness, and Eigenvector Centralities

	Southern Women N = 18				
	Depend.	Degree	Close.	Between.	Eigen.
Depend.	1	0.97258	0.76232	0.85937	0.95036
Degree		1	0.70454	0.84009	0.99327
Close.			1	0.90882	0.66326
Between.				1	0.81578
Eigen.					1
	al-Qaida Members N = 18				
	Depend.	Degree	Close.	Between.	Eigen.
Depend.	1	0.81706	0.85895	0.85391	0.67149
Degree		1	0.95024	0.85663	0.96572
Close.			1	0.91478	0.87978
Between.				1	0.75056
Eigen.					1
	Facebook-like Forums N = 899				
	Depend.	Degree	Close.	Between.	Eigen.
Depend.	1	0.92533	0.89331	0.92518	0.91228
Degree		1	0.96826	0.90876	0.97648
Close.			1	0.90624	0.98598
Between.				1	0.92203
Eigen.					1
	Mean Across Test Networks				
	Depend.	Degree	Close.	Between.	Eigen.*
Depend.	1	0.90499	0.83820	0.87949	0.84471
Degree		1	0.87435	0.86849	0.97849
Close.			1	0.90995	0.84301
Between.				1	0.82946
Eigen.					1

** This mean treats the undirected eigenvector centrality scores and the directed Page centrality scores as comparable.*

6. Conclusions

This paper summarized a flow-based method to transform bipartite data into one-mode networks that was previously described by Zhou et al. This paper also presented a generalization that allows this method to be applied to weighted networks. However, this paper's primary contribution is to present dependency centrality, a new measure that the existing literature on centrality does not appear to describe.

This measure performed differently than widely-implemented measures of centrality across three test networks. Thus, dependency centrality appears to estimate aspects of structural leadership that are overlooked by existing means of determining node-level importance.

As the measure's name implies, dependency centrality more accurately scores the extent to which nodes in a bipartite network are structurally reliant on high-ranking nodes. For agent x event networks, high dependency centrality scores suggest that an agent has a large share of individuals who participate in events only alongside the central actor, and low dependency centrality scores that are greater than zero suggest that an agent tends not to participate in events absent more central individual(s). In the unique case that an agent is not an isolate in the original bipartite network, but has a dependency centrality of zero, it indicates that the agent is a lone wolf, who neither follows a central leader nor possesses adherents of his own. Thus, dependency centrality appears to capture the dynamics of mentormentee, recommender-recommendee, and other similar relationships.

However, leadership and importance are diffuse concepts, the definitions of which often change based on context, and diverse measures are necessary to assess different aspects of these concepts in different analytic environments. Dependency centrality is a reasonably specialized measurement that only functions in bipartite networks and cannot be applied in 'natural' one-mode networks. Moreover, Kleinberg's discussion of the features of online networks, which are replete with dubious connections and contain nodes that have limitless capacity to acquire alters, strongly suggests that dependency centrality should only be implemented in social environments and other contexts in which nodes in the bipartite graph have limited ability to form connections. While dependency centrality ultimately lacks the broader generalizability of degree, closeness, betweenness, and eigenvector-based measures of centrality, the new measure provides a distinct and valid means to assess node-level importance that estimates the extent to which nodes are structurally reliant on one another. Dependency centrality has clear utility for analysts attempting to determine organizational structure, analysts studying recommendation patterns among consumers, and analysts examining other socially-oriented issues that feature a leader-follower dynamic.

References

Bjelland, J., Burgess, M., Canright, G., & Engø-Monsen, K. (2010). Eigenvectors of directed graphs and importance scores: dominance, T-Rank, and sink remedies, *Data Mining and Knowledge Discovery*, Vol. 20, pp. 98-151.

Bolland, J.M. (1988). Sorting out centrality: An analysis of the performance of four centrality models in real and simulated networks. *Social Networks*, Vol. 10, No. 3, pp. 233-253.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Mathematical Sociology*, Vol. 2, No. 1, pp. 113-120.

Borgatti, S.P., & Everett, M.G. (1997). Network analysis of 2-mode data. *Social Networks*, Vol. 19, pp. 243-269.

Breiger, R. (1974). The duality of persons and groups. *Social Forces*, Vol. 53, pp. 181-190.

Burt, R.S. (1987). Social contagion and innovation: Cohesion versus structural equivalence. *American Journal of Sociology*, Vol. 92, pp. 1287-1335.

Center for Computational Analysis of Social and Organizational Systems (CASOS). (2008). Tanzania-Kenya-imoon.xml. Data available online: http://www.casos.cs.cmu.edu/computational_tools/datasets/internal/tanzania_kenya/index11.php.

Coleman, J.S., Katz, E., & Menzel, H. (1966). *Medical Innovation: A Diffusion Study*. New York, NY: Bobbs Merrill.

Csardi, G. & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal, Complex Systems*, Vol. 1695.

Davis, A., Gardner, B.B., & Gardner, M.R. (1941). Deep South: A Social Anthropological Study of Caste and Class.. Chicago, IL: University of Chicago Press. Data available online: http://members.ozemail.com.au/~dekker/bipartite/.

Faust, K. (1997). Centrality in affiliation networks. *Social Networks*, Vol. 19, pp. 157-191.

Freeman, L.C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, Vol. 1, pp. 215-239.

Freeman, L.C. (2004). *The Development of Social Network Analysis: A Study in the Sociology of Science*. Vancouver, BC: Empirical Press, pp. 18-19.

Gerdes, L.M. (2008). Codebook for network data on individuals involved with terrorism and counterterrorism. *Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report CMU-ISR-08-136*. Available online: http://www.casos.cs.cmu.edu/publications/papers.php.

Gerdes, L.M. (2014). MAPPing dark networks: A data transformation method to study clandestine organizations. *Network Science*, Vol. 2, No. 2, pp. 213-253.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM*, Vol. 46, No. 5, pp. 604-632.

Latapy, M., Magnien, C., & Del Vecchio, N., (2008). Basic notations for the analysis of large two-mode networks. *Social Networks*, Vol. 30, No. 1, pp. 31-48.

Lee, C. (2006). Correlations among centrality measures in complex networks. *Physics and Society*.

Li, M., Dias, B., Jarman, I., El-Deredy, W., & Lisboa, P. (2009). Grocery shopping recommendations based on basket-sensitive random walk. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY:

Association for Computing Machinery, pp. 1215-1224.

Liu, J., Shang, M., & Chen, D. (2009). Personal recommendation based on weighted bipartite networks. *Sixth International Conference on Fuzzy Systems and Knowledge Discovery*.

Lorrain, F., & White, H. (1971). Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, Vol. 1, pp. 49-80.

McDonald, J.H. (2009a). Spearman rank correlation. *Handbook of Biological Statistics*, 2nd ed. Baltimore, MD: Sparky House Publishing; available online: http://udel.edu/~mcdonald/statspearman.html.

McDonald, J.H. (2009b). Multiple comparison. *Handbook of Biological Statistics*, 2nd ed. Baltimore, MD: Sparky House Publishing; available online: http://udel.edu/~mcdonald/statmultcomp.html.

Newman, M.E.J. (2001). Scientific collaboration networks. II. Shortest Paths, weighted networks, and centrality. *Physical Review E*, Vol. 64, 016132.

Newman, M.E.J. (2004). Analysis of weighted networks. *Physical Review E*, Vol. 70, 056131.

Opsahl, T. (2009). *Structure and equivalence of weighted networks*. London: UK: University of London, Queen Mary College, pp. 104-122; available online: http://toreopsahl.com/publications/thesis/ or http://toreopsahl.com/tnet/.

Opsahl, T. (2013). Triadic closure in two-mode networks: Redefining the global and local clustering coefficients. *Social Networks*, Vol. 35, No. 2, pp. 159-167. Data available online: http://toreopsahl.com/datasets/#online_forum_network.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). The pagerank citation ranking: Bringing order to the web. *Technical report, Stanford Digital Library Technologies*.

Padrón, B., Nogales, M., & Traveset, A. (2011). Alternative approaches of transforming bimodal into unimodal mutualistic networks: The usefulness of preserving weighted information. *Basic and Applied Ecology*, Vol. 12, pp. 713-721.

Pan, Y., Li, D., and Liang, J. (2010). Detecting community structure in complex networks via node similarity. *Physica A: Statistical Mechanics and its Applications*, Vol. 389, No. 14, pp. 2849-2857.

Ramsey, P.H. (1989). Critical values for Spearman's rank order correlation. *Journal of Educational Statistics*, Vol. 14, No. 3, pp. 243-253.

Valente, T.W. & Foreman, R.K. (1998). Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Social Networks*, Vol. 20, pp. 89-109.

Valente, T.W., Coronges, K., Lakon, C., & Costenbader, E. (2008). How correlated are network centrality measures? *Connections*, Vol. 28, No. 1, pp. 16-26.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York, NY: Cambridge University Press.

Zhou, T., Ren, J., Medo, M., & Zang, Y. (2007). Bipartite network projection and personal recommendation. *Physical Review E*, Vol. 76, 046115.

The Operating Room: It's a Small World (and Scale Free Network) After All

Patrick J. Tighe
University of Florida College of Medicine
Gainesville, Florida

Laurie Davies
University of Florida College of Medicine
Gainesville, Florida

Stephalie Stephanie Patel
University of Florida College of Medicine
Gainesville, Florida

Stephen D. Lucas
University of Florida College of Medicine
Gainesville, Florida

Nikolaus Gravenstein
University of Florida College of Medicine
Gainesville, Florida

H. Russell Bernard
University of Florida College of Medicine
Gainesville, Florida

Abstract
Introduction: An operating room's (OR) organizational behavior, including its susceptibility to certain types of failure, may partially reflect its structural features. We report the results of a structural analysis of a composite OR suite in a tertiary-care teaching hospital. Methods: We conducted a simulation study of the OR interaction network in a 900-bed teaching hospital. A composite OR network was built from a single-day operating room schedule encompassing 32 anesthetizing sites. There were two aims: (1) to compare the composite, or prototypical, OR network to three network types: random, scale-free, and small-world; (2) to calculate the total degree centrality, eigenvector centrality, and betweenness centrality for each node within the prototypical OR network, and to compare these metrics by level of physician training and by OR role. Results: The complete prototypical OR network included 146 nodes linked by 329 edges. Results indicate that the OR is a scale-free network with small-world characteristics. The chief anesthesiologist, OR charge nurse, and recovery room charge nurse had the highest total degree centralities. There were significant differences in total degree centrality scores between nurses and anesthesiologists, nurses and surgeons, and anesthesiologists and surgeons; attending physicians had greater perioperative total degree centrality than did resident physicians. Conclusion: Given the homogeneity of certain scale-free network characteristics throughout nature, such a designation has potentially critical implications for coordinating anesthesiologists and nurses, whose roles will be impacted by the continued growth of operating rooms. These implications will be tested at the next stage of the project.

Authors
Patrick J. Tighe is an Assistant Professor of Anesthesiology at the University of Florida College of Medicine in Gainesville, Florida.

Sephalie Stephanie Patel is a Resident in Anesthesiology at the University of Florida College of Medicine in Gainesville, Florida.

Nikolaus Gravenstein is a Professor of Anesthesiology at the University of Florida College of Medicine in Gainesville, Florida.

Laurie Davies is an Associate Professor of Anesthesiology at the University of Florida College of Medicine in Gainesville, Florida.

Stephen Lucas is an Assistant Professor of Anesthesiology at the University of Florida College of Medicine in Gainesville, Florida.

H. Russell Bernard is a Professor of Anthropology, emeritus, at the University of Florida College of Liberal Arts and Sciences in Gainesville, Florida.

Please send all correspondence to: Patrick J. Tighe, MD, MS, Department of Anesthesiology, University of Florida College of Medicine, PO Box 100254, Gainesville, FL 32610-0254. Telephone: 352-294-5076; Fax: 352-265-8013; e-mail: ptighe@anest.ufl.edu

Acknowledgements
Funding provided by NIH K23 GM102697 and the Jerome H. Modell, MD, Endowed Professorship

---

1. Introduction

"They just got into the vena cava. Can you please send someone to help hang blood?"

"Anesthesia, charge nurse stat to OR 5….Anesthesia, charge nurse stat to OR 5…."

"I have to pee. Can I get a break out?"

"The patient's tumor is unresectable. We'll be closing here shortly."

"Trauma, coming up now."

"I'm calling to post a heart-lung transplant. Should start around midnight."

"I think my water just broke. Don't think I'll be able to finish this case."

"I'm calling to post an appendectomy to OR 7…"

Complexity confounds the modern operating room theater. The above eight phone calls offer a glimpse into the incessant communiqués assailing charge nurses and chief anesthesiologists in many hospitals around the world. Responses are often necessary within seconds or minutes. Further complicating matters, these phone calls can arrive in a span of minutes. Even brief periods of unavailability by key personnel can lead to trouble as surgical staff must spend valuable seconds looking up second- and third-tier contact information. Such challenges have led many hospitals to designate specific anesthesiologists and nurses, each equipped with multiple cell phones and beepers, whose role is simply to coordinate the perioperative complexity at hand.

Such complexity is frequently cited as a significant hurdle to improving perioperative patient safety (ElBardissi & Sundt, 2012). This complexity may be considered at three levels of analysis: micro (pertaining to small teams of individuals or conceptual entities), meso (pertaining to organizational scales), and macro (pertaining to large populations). Efforts to reduce the number of perioperative harm events typically focus on actions such as antibiotic timing, temperature regulation, deep vein thrombosis prophylaxis, and more recently, aggregating these micro-level actions through the use of checklists or bundled measures (Haynes, Weiser, Berry, Lipsitz, Breizat, Dellinger, et al. 2009; Stulberg, Delaney, Neuhauser, Aron, Fu, & Koroukian, 2010).

Prior work exploring operating room (OR) social networks has mainly focused on micro-level networks, as contrasted with the meso-level network scale investigated here (Chamers, Wilson, Thompson, & Harden, 2012). Anderson and Talsma (2011) demonstrated that team coreness was associated with surgical duration, with cases scheduled earlier in the day demonstrating high coreness of surgical teams. Perhaps not surprisingly, as individuals completed their shifts at disparate times, team coreness diminished throughout the day, a change that was associated with longer surgical duration. This theme was reinforced by the findings of Listyowardojo (2012), whereby more complex surgical procedures were accompanied by increasingly interrelated subgroups of OR team members. Baumgart, Denz, Bender & Schleppers (2009) studied social network analysis effects during the planning phases of an ambulatory surgical center redesign following the physical rearrangement of ORs from a dispersed to a centralized OR layout. They noted that this intervention, while reducing the risk of deviations from standard care practices, nevertheless introduced greater interdependency among staff with increasing errors during handovers. Social network analysis metrics have also been shown to be directly associated with patient outcomes such as adverse drug events and patient falls, although these findings have yet to be applied to perioperative patient populations (Effken, Carley, Gephart, Verran, Banchi, Reminga, & Brewer, 2011).

Less is known about meso- and macro-level factors that affect patient safety, but evidence has been building for some time that the structure of meso-scale interactions in a group affects outcomes in epidemiology, defense, logistics, and social media (e.g., Facebook and Twitter) and that these effects can be managed to improve work efficiency (Krackhardt, 1990; Krackhardt & Hanson, 1993; Krebs, 2002; Nakano & White, 2006;

Small & Chi, 2005; Xu, Hu, & Chen, 2009).

The long-term goal of our project is to test whether the structural features of three major network types—random, small-world, scale-free—are associated with perioperative risks in a large, tertiary care teaching hospital. For example, are various OR roles, especially those that are most sensitive to failure, associated with particular measures of centrality? In the current phase of our project, we model the answers to these questions and establish a platform for direct observation of the OR and for analysis at both the micro- and meso-scales of organizational complexity.

2. Methods

Institutional review board approval was not required because this was a simulation study. We conducted a simulation study that involved organizational network analysis of the OR social networks within a 900 bed teaching hospital (UF Health at the University of Florida, Gainesville, Florida).

Description of Operating Room Environment

For each surgery, a patient is cared for by an OR nurse, an anesthesiologist, and a surgeon. In an academic teaching hospital, the anesthesiologist is frequently a resident (e.g, trainee) physician, supervised by a more senior and board-certified attending anesthesiologist. Each attending anesthesiologist may supervise one or two residents. Alternatively, the attending anesthesiologist may supervise up to three or four certified registered nurse anesthetists (CNRAs) or anesthesiology assistants (AAs). The number of residents, CRNAs or AAs supervised depends both upon the type of surgery and medical complexity of the patients. Similarly, the attending surgeon may supervise one or two resident surgeons, again depending upon the type of procedure and medical complexity of the patient. Such divisions of labor are frequently typified; for instance, a general surgeon may supervise two resident surgeons, one performing a straightforward appendectomy and the other a repair of a simple hernia. On the other hand, replacement of a heart valve on cardiopulmonary bypass is a procedure involving intense and continual resident supervision.

The lines of communication amongst these team members can also be typified. For instance, the attending surgeon frequently discusses the progress of the surgery with the anesthesiologist and nurse in the OR. The resident anesthesiologist will update his or her attending physician at regular intervals, and the nurse will relate changes to his or her charge nurse. Continuing with this example, the resident anesthesiologist will typically not need to initiate a discussion with a surgeon in another OR, nor will the circulating nurse need to initiate conversation with the lead anesthesiologist of the day.

The phone numbers for each OR's attending anesthesiologist, surgeon, nurse, charge nurse, and anesthesiologist of the day are made readily available by writing them on a large white board within each OR. Further, the charge nurse and anesthesiologist of the day each carry a cellular telephone dedicated to these respective roles. Most staff in an OR have memorized these two telephone numbers. The contact information for all other anesthesiologists, surgeons, and nurses is available in a large binder in the OR and through a computerized database. However, the time and effort required to isolate a given individual's contact information from these two sources, while often incremental over that of the aforementioned whiteboard or memory, become nearly insurmountable in times of urgency or emergency when all hands within the OR may be dedicated to more critical tasks.

For this study, we produced a composite OR network from data on 32 anesthetizing sites extracted from a single day of operating room schedules. The network was created by comparing direct observations, case assignments, and protocol-based hierarchical reporting lines within the OR . We also conducted multiple interviews with three attending anesthesiologists (authors NG, SL, and LD), each with over 10 years of experience at this institution and regular assignments as the coordinating anesthesiologist for the ORs.

At each stage of the network's construction, the draft network model was reviewed using a modified Delphi method (Dalkey & Helmer, 1963) until a consensus model was obtained. At each revision, interviewees were asked to constrain their design to that of an "average weekday" based upon their own personal experience and observation. The intent was for the model to capture the interpersonal interactions of an "average weekday" over a 12-hour timeframe. For this process, we first created a prototype micro-level network of an OR "team" included circulating nurses, anesthesiology residents, AAs, CRNAs, anesthesiology attendings, and surgeons. These teams were replicated and serially adjusted to reflect different surgical specialties and their daily room allocation at the hospital. The teams were then connected according to anesthesiology attending assignment, reporting to charge nurses, and reporting to the anesthesiologist of the day, according to the formal reporting structure at our institution.

Communications Network for Composite Prototypical Operating Room

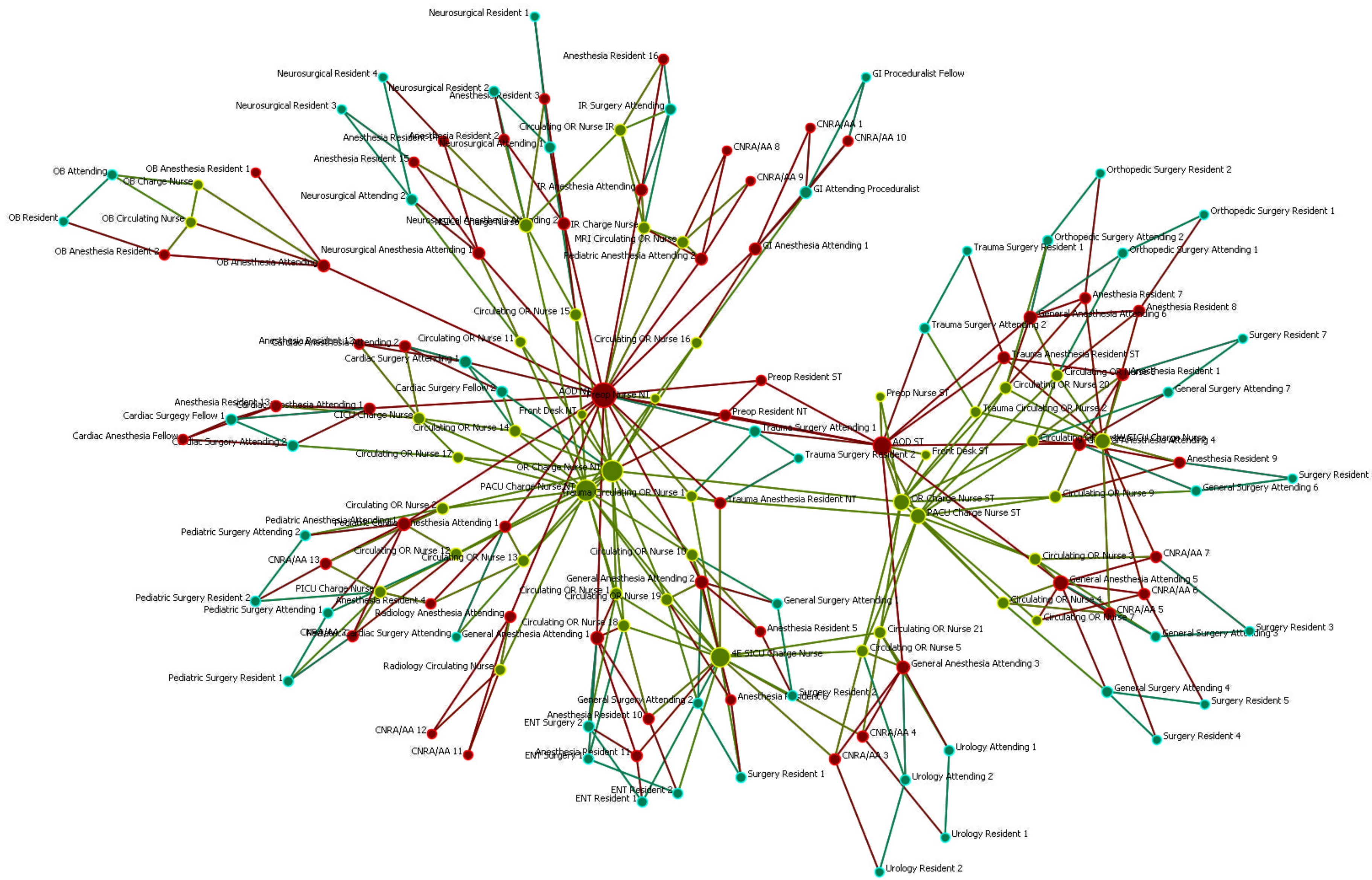


Figure 1: Plot of the Operating Room Communication Network

Table 1: Centraility Scores for High-Ranking Operating Room Nodes

Rank	Total Degree Centrality		Eigenvector Centrality		Betweenness Centrality	
	Agent	Score	Agent	Score	Agent	Score
1	AOD NT	24	OR Charge Nurse NT	0.38	AOD NT	4470.06
2	PACU Charge Nurse NT	19	PACU Charge Nurse NT	0.36	PACU Charge Nurse NT	2483.81
3	OR Charge Nurse NT	19	AOD NT	0.31	AOD ST	2438.00
4	4E SICU Charge Nurse	17	4E SICU Charge Nurse	0.24	PACU Charge Nurse ST	1449.32
5	AOD ST	14	AOD ST	0.17	OR Charge Nurse NT	1259.23
6	OR Charge Nurse ST	12	Circulating OR Nurse 18	0.16	4E SICU Charge Nurse	1163.86
7	PACU Charge Nurse ST	11	Circulating OR Nurse 10	0.16	OR Charge Nurse ST	931.64
8	4W SICU Charge Nurse	11	Circulating OR Nurse 19	0.16	OB Anesthesia Attending	839.50
9	General Anesthesia Attending 5	9	Circulating OR Nurse 1	0.15	NSICU Charge Nurse	569.34
10	NSICU Charge Nurse	8	OR Charge Nurse ST	0.15	General Anesthesia Attending 5	538.34
11	General Anesthesia Attending 1	7	PACU Charge Nurse ST	0.14	GI Anesthesia Attending 1	404.44
12	General Anesthesia Attending 2	7	Circulating OR Nurse 14	0.13	General Anesthesia Attending 6	402.27
13	Pediatric Anesthesia Attending 1	7	Preop Resident NT	0.13	Pediatric Anesthesia Attending 1	352.12
14	General Anesthesia Attending 3	7	Circulating OR Nurse 15	0.12	General Anesthesia Attending 4	340.12
15	General Anesthesia Attending 4	7	Circulating OR Nurse 12	0.12	Neurosurgical Anesthesia Attending 1	338.07
16	General Anesthesia Attending 6	7	Circulating OR Nurse 2	0.12	Cardiac Anesthesia Attending 1	319.61
17	CICU Charge Nurse	6	General Anesthesia Attending 1	0.12	Neurosurgical Anesthesia Attending 2	300.61
18	Circulating OR Nurse 10	6	Circulating OR Nurse 13	0.12	General Anesthesia Attending 2	299.94
19	Circulating OR Nurse 18	6	Circulating OR Nurse 17	0.12	PICU Charge Nurse	296.01
20	Circulating OR Nurse 19	6	Circulating OR Nurse 16	0.11	General Anesthesia Attending 1	284.08

Description of Nodes & Edges

Each node in the network (Figure 1) represents an individual healthcare provider involved with direct patient care. Nodes were classified as attending anesthesiologists, attending surgeons, anesthesiology residents and fellows, CRNAs or AAs, surgical residents and fellows, OR circulating nurses, and OR charge nurses. Each node was weighted equally, although physician nodes were stratified by training status (resident/fellow versus attending physician). The edges connecting each node were weighted equally at 1 and all were considered bidirectional. Edges consisted of direct or telephone contact commonly used for perioperative communications.

Analysis

Out first goal was to compare the OR network to three major network types: random, scale-free, and small-world. The node and link count, clustering coefficient, and average path length were first calculated for the whole network. In comparing the structure of the network in Figure 1 with a random network, nodal degree centrality distribution was tested against the Poisson distribution using the Pearson chi-square goodness-of-fit test (Barabási, 2009). For small-world network comparison, the clustering coefficient and mean shortest path length were calculated for the OR network and for a randomized version of the network. The randomized version of the OR network used the same nodes as in the native network, but the links between the nodes were randomly distributed according to the method of Erdős & Rényi (1959). Quantification of small-worldness was calculated according to the method of Humphries & Gurney (2008), relying on a quantification of strong clustering coefficients but low mean shortest path lengths. For comparison with a scale-free network, we performed a Kolmogorov-Smirnov (KS) goodness-of-fit test, modified for power-law distributions of degree centrality, according to the method of Clauset, Shalizi & Newman (2009). Further details concerning calculation of the goodness-of-fit to the power law distribution are given in Appendix A.

Our second goal was to calculate total degree centrality, eigenvector centrality, and betweenness centrality (Bonacich, 1972; Freeman, 1978) for each node in the network and then compare the top 20 scores across nurses, surgeons, and anesthesiologists. Further comparisons were conducted between trainees and attendings for surgeons and anesthesiologists. Operating room circulator nurses, AAs, and CRNAs were excluded from the analysis of training status.

The network analysis was performed using Organizational Risk Analyzer 2.0.04 (Center for Computational Analysis of Social and Organizational Systems (CASOS), Carnegie Mellon University, Pittsburgh, PA). Comparisons were conducted using a non-parametric method (Kruskal-Wallis test) to compare centrality scores of anesthesia, surgery, and nursing roles, with post-hoc comparisons corrected using the Steel-Dwass method. Comparisons of trainee versus attending physician centrality scores were conducted using the Wilcoxon test. A pre-study power analysis was not conducted due to the empiric nature of the network design. Alpha was set to 0.05. All statistical comparisons were conducted using R 2.15.2 (http://www.r-project.org).

4. Results

Network Composition

The complete OR network included 146 nodes linked by 329 edges (Figure 1). The clustering coefficient of the OR network was 0.326 with a diameter of 7 and fragmentation of zero. Mean shortest path length was 3.685.

Comparison of the OR Network with Prototypical Networks

Distribution of total degree centrality did not follow a Poisson distribution pattern (Pearson chi square 7.82 x 10^7p<0.0001), suggesting that the OR network was not a prototypical random network. The clustering coefficient for the randomized OR network was 0.041, and mean shortest path length was 2.718, leading to an S ratio of 5.864 and thus denoting a Watts-Strogatz small-world network characteristic (Watts & Strogatz, 1998).

When compared to a power-law distribution, an x_{min} of 4 out of 146 nodes lead to a KS statistic of D=0.041 (p=0.6) with $\alpha = 3.4$, thus confirming a null hypothesis that the total centrality degree distribution fit a power law distribution. Goodness-of-fit tests against the exponential distribution (KS statistic D=0.35, p<0.01) and log-normal distribution (KS statistic D=0.144, p<0.01) further suggest that neither of these alternative distributions surpasses the fit provided by the power law distribution for the observed total degree centrality scores (Figure 2).

Nodal Metrics

The numbers and proportions of OR team members according to role and training status are given in Figure 3. The anesthesiologist of the day for the North Tower (AOD NT), who coordinates daily staffing; the PACU charge nurse of the NT; and the OR charge nurse NT scored the highest total degree centrality with scores of 24, 19, and 19 (Table 1). The OR charge nurse NT and PACU charge nurse NT had the highest eigenvector centralities of 0.38 and 0.36, followed by the AOD NT with a score of 0.31. For betweenness centrality, the AOD NT had the highest score of 4470.1, followed by the PACU charge nurse NT (2483.8) and anesthesiologist of the day for the South Tower (AOD ST) (2438).

Comparisons of Centrality Scores by Perioperative Role

Total degree centrality scores differed significantly by OR role (Figure 4). Nursing had the highest total degree centrality scores (median 5, range 2-19), although anesthesia had the largest range (median 4, range 1-24). Surgery had both the lowest median and range (median 3, range 2-5) for total degree centrality. Eigenvector centrality scores showed statistically significant differences among all combinations involving nursing (median 0.092, range 0.0076-0.38), anesthesia (median 0.041, range 0.0061-0.31), and surgery (median 0.019, range 0.0013-0.085). Anesthesia (median 83.8, range 0-4470.1) and nursing (median 157, range 0-2483.9) each had greater betweenness centrality scores than did surgery (median 8.8, range 0-170.9).

Comparisons of Centrality Scores by Level of Training

The scores for anesthesia and surgical attending physicians were greater than those for residents for total degree centrality (attending median 5, range 2-24; resident median 3, range 1-5; p<0.0001), eigenvector centrality (attending median 0.044, range 0.0023-0.31; resident median 0.019, range 0.0013-0.13; p<0.0001), and betweenness centrality (attending median 131.7, range 0-4,470.1; resident median 5, range 0-192.5; p<0.0001) (Figure 3).

5. Discussion

Analysis shows that the organizational structure of our prototypical OR is consistent with a scale-free network, rather than a random network (Barbasi & Albert, 1999; Klem & Eguíluz, 2002a; Klem & Eguíluz, 2002b). It also contains features of a small-world network in that

Simulated Erdos-Renyi Prototype of Operating Room Communications Network

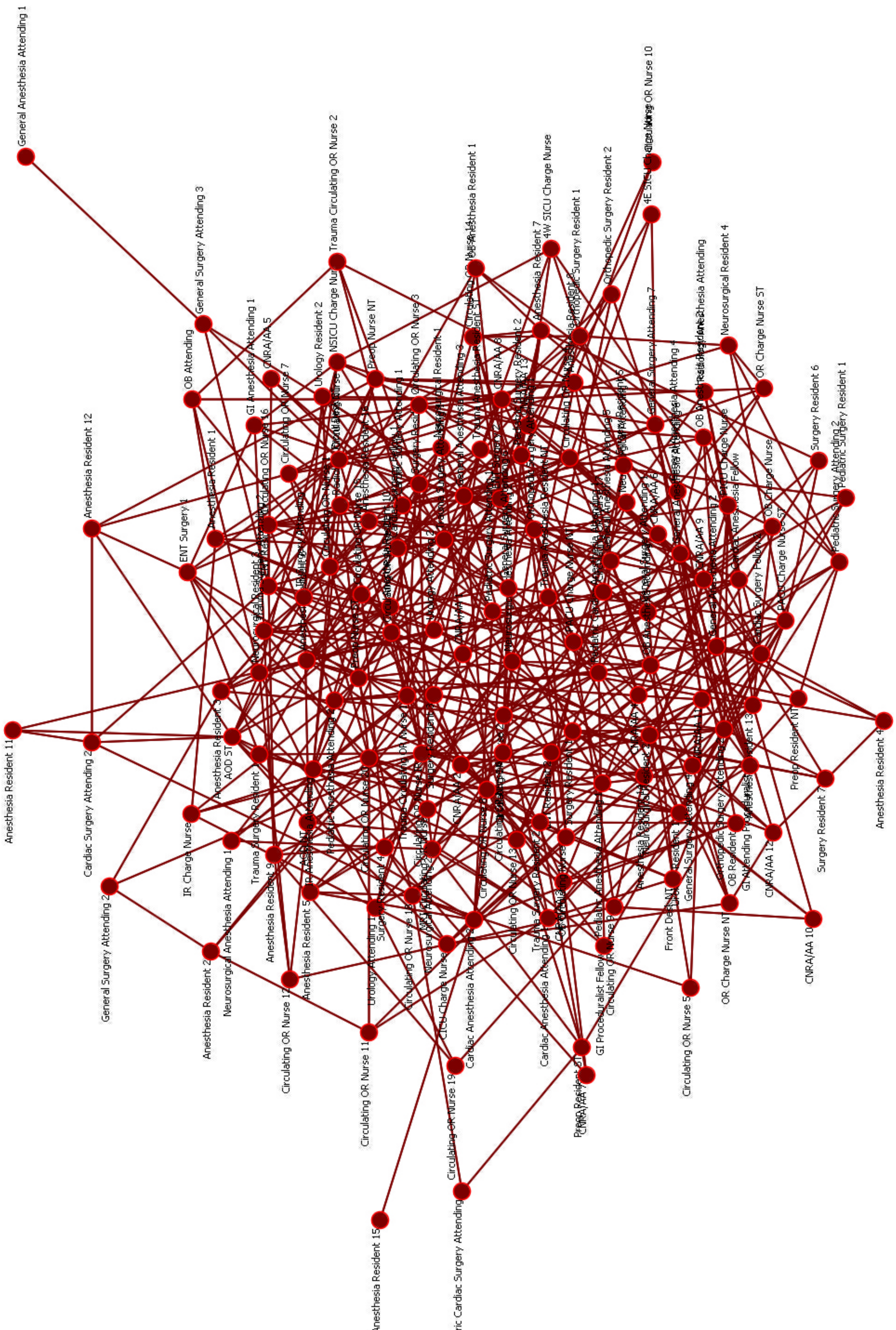


Figure 2a: Proportions of OR Network Healthcare Providers by Role and Training Status

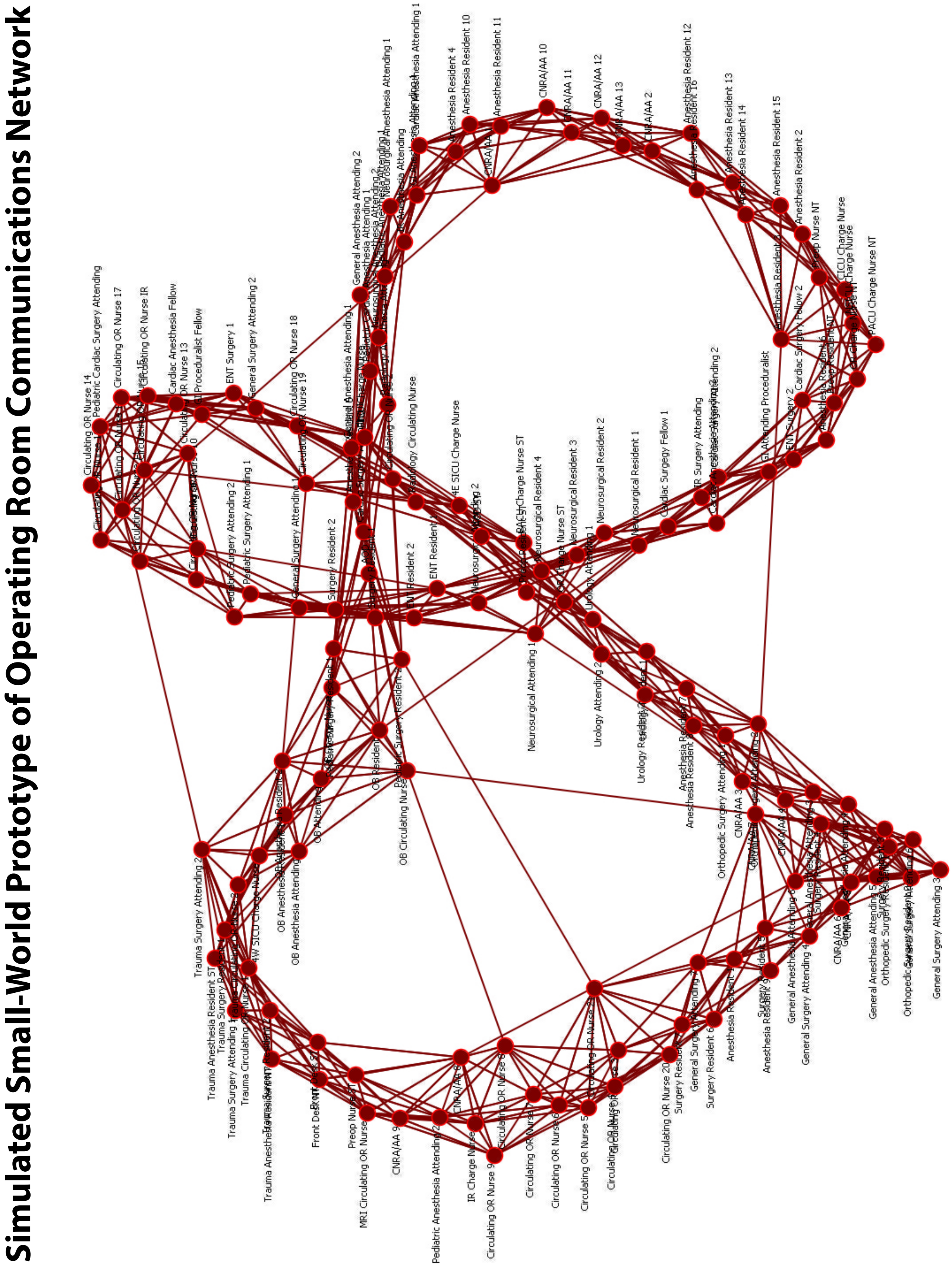


Figure 2b: Example Prototypes and Small-World Networks

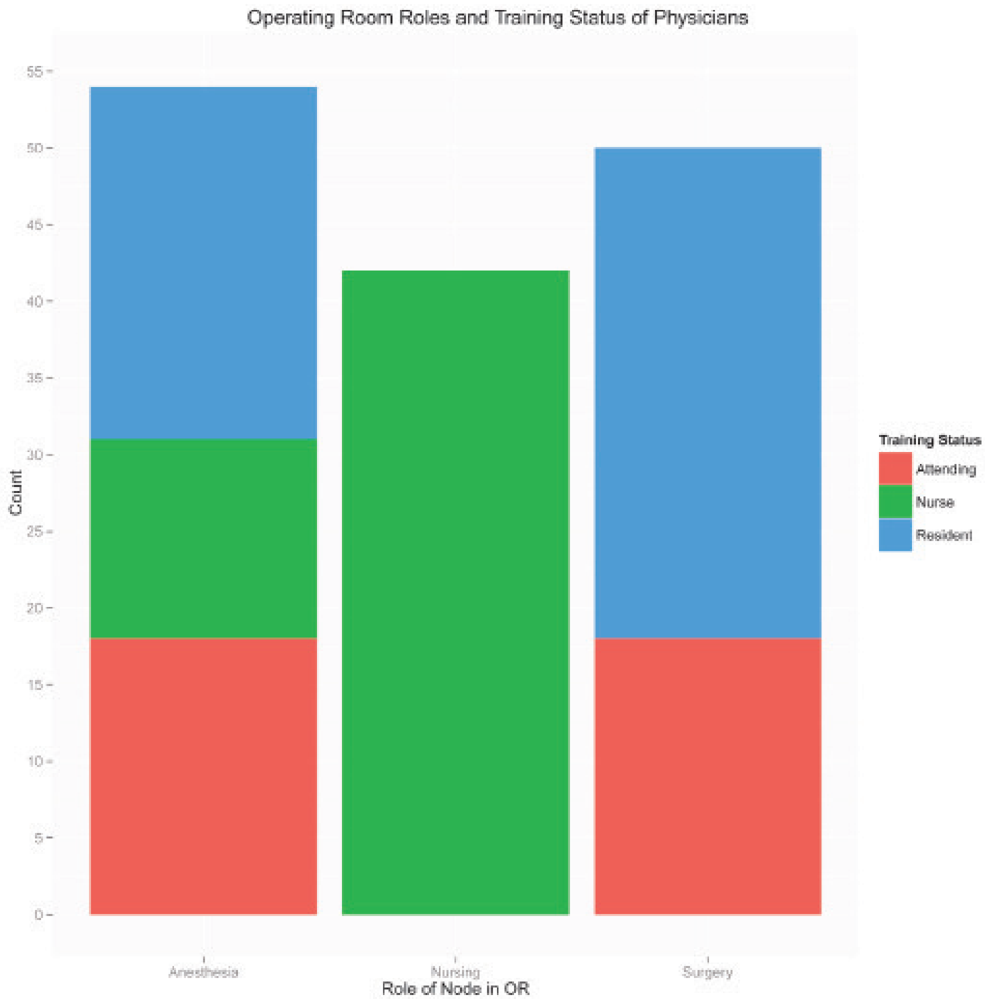


Figure 3: Centraility Scores for High-Ranking Operating Room Nodes

disparate nodes, located far from each other across the network, are closely related through well-connected nodes. The number of such well-connected nodes in a scale-free network remains low in proportion to the number of nodes within said network. Senior anesthesia and nursing staff (but not surgical staff in the studied hospital where the schedule is managed by a chief anesthesiologist, shown in Figure 1 as the AOD), are the most well-connected on three centrality scores, including total degree, eigenvector, and betweenness. Thus, both anesthesiology staff and nurses are connected to everyone else in the graph through a small number of extremely well-connected nodes that are also directly connected. This connectivity facilitates the small-world effect by increasing clustering and decreasing path lengths. As expected, surgical and anesthesia attending physicians consistently outranked resident physicians across all measures of centrality because attending physicians carry greater responsibility than trainees.

Network behaviors are frequently influenced by the underlying network structure type, of which multiple prototypes have been identified throughout nature (Albert & Barabasi, 2002). One of the original network structures described is the random network of Erdős, & Rényi (1959), where an edge (i.e, interpersonal interaction) can be found between any given pair of nodes with equal probability, independent of the presence of other edges in the network. Small-world networks, of which it appears the OR is an example, represent a class of networks in which: (1) most nodes are distant from one another, and (2) most can be reached by other nodes through a small number of connections to highly connected nodes.

A third network prototype, of which our OR is also an example, is the Barabasi-Albert model of the scale-free network. Here, the distribution of node centrality follows a power-law distribution (Albert, Jeong, & Barabasi, 2000; Barabasi & Albert, 1999). Scale-free networks exhibit three features with important implications for OR organization:

1. Scale-free, or power-law, distributions of total degree centrality. A small number of "hub" individuals—in the case of the OR, the coordinating anesthesiologist, and charge nurses—contain a disproportionately large

share of the total number of connections within the OR network. This is in contrast to random networks, whereby each individual is likely to maintain a similar number of connections to other individuals.

2. Preferential attachment behaviors, such that the greater centrality of a given node, the greater the chance that it will receive additional connections to other nodes. Applied to an OR, this suggests that well-connected nodes such as the AOD and OR charge nurse will continue to attract connections to new nodes at a rate greater than that of other members of the OR team as the OR network expands. Most scale-free networks also exhibit consistent growth rather than static structure (Barabasi, 2009). This growth is thus expected to disproportionately affect the AODs and charge nurses who already bear a disproportionately large number of connections. Prior observational studies of communication burdens in the emergency department and operating room (Woloshynowych, Davis, Brown & Vincent, 2007; Hu, Arriaga, Peyre, Corso, Roth, & Greenberg, 2012) lead us to predict that the expansion of the OR network will place these individuals at increasing risk of communication errors due to mismatches between demand and capacity. Importantly, systems based on scale-free networks are robust to stochastic failures across the network but sensitive to focused "attacks" on a small number of critical nodes (Cohen, Erez, ben-Avraham, & Havlin, 2000). Although communication errors of most individuals within the OR may affect only local members of the network, errors by the AOD or charge nurses can lead to catastrophic failures, that reverberate throughout the entire network. Given the characteristics of preferential attachment and growth inherent to scale-free networks, this suggests that those to whom the network is most sensitive to failure, in our case the AOD and charge nurse, will continue to be excessively burdened with additional connections—and at an ever-increasing rate.

Given the costs associated with medical errors, and the fact that communication errors contribute to the vast majority of medical errors, there is a current emphasis on improving communication within the ORs (Eichorn, 2013). In a scale-free network with small-world properties, protecting a few highly-connected individuals may be more prudent. Indeed, epidemiologic models incorporating small-world and scale-free features suggest that such features place even greater importance in protecting key entities within critical, dynamic environments (Small & Chi, 2005). Another approach may consider placement of senior staff into roles based not just on high patient acuity, but also in consideration of roles requiring critical communication volume with large numbers of OR team members. Dissemination of practice updates may also focus on those highly connected individuals because those individuals will carry not only a high level of present-tense connectivity, but are also well positioned to disseminate information to new nodes given the principle of preferential attachment within scale-free networks (Anderson & Jay, 1985; Coleman, Katz, & Menzel, 1957). Together, these proposals will require further investigation to ascertain whether their implementation is both practical and efficacious.

We also examined the relationship between OR role and nodal metrics. The nursing staff had higher scores in all centrality metrics, with anesthesiology staff similar to nursing, and surgical staff significantly lower than both nursing and anesthesiology. Several nurses (specifically, the charge nurses) are tracking multiple ORs, leading them to be the most directly connected in the network. Anesthesiology had similar scores to nursing because most anesthesia attendings were running more than one operating room; the AOD NT and AOD ST had connections to the entire OR-NT and OR-ST, and the preoperative residents had multiple connections as well. This organizational connectivity, is what allowed anesthesia to have higher total degree and eigenvector centrality scores. As expected, surgeons were lower in all scores because each attending surgeon is typically focused on the events in his or her single OR.

Although our work provides insight into the meso-scale organizational structure of the OR, many additional facets warrant exploration. At the micro-level, structural features of individual networks (personal networks) may impact perioperative patient outcomes. Prior clinical research has often focused on an individual's contribution to outcome variance, perhaps most famously in reference to anesthesiologist number 7 (Slogoff & Keats, 1985).

The story of anesthesiologist number 7 has attained legendary status among academic anesthesiologists. Slogoff and Keats (1985) attempted to determine whether myocardial ischemia during surgery to bypass coronary vessels was associated with myocardial infarctions following surgery. During their study, it was incidentally found that patients of one particular anesthesiologist, anesthesiologist number 7, had the highest rates of perioperative myocardial ischemia and infarction, likely due to the very fast heart rates induced by this individual. "Anesthesiologist No. 7," the authors reiterated in a follow-up on classic papers in *Anesthesiology*, "dramatically demonstrated the importance of preventing and treating tachycardia (fast heart rate) in the population with coronary artery disease" (Slogoff & Keats, 2006).

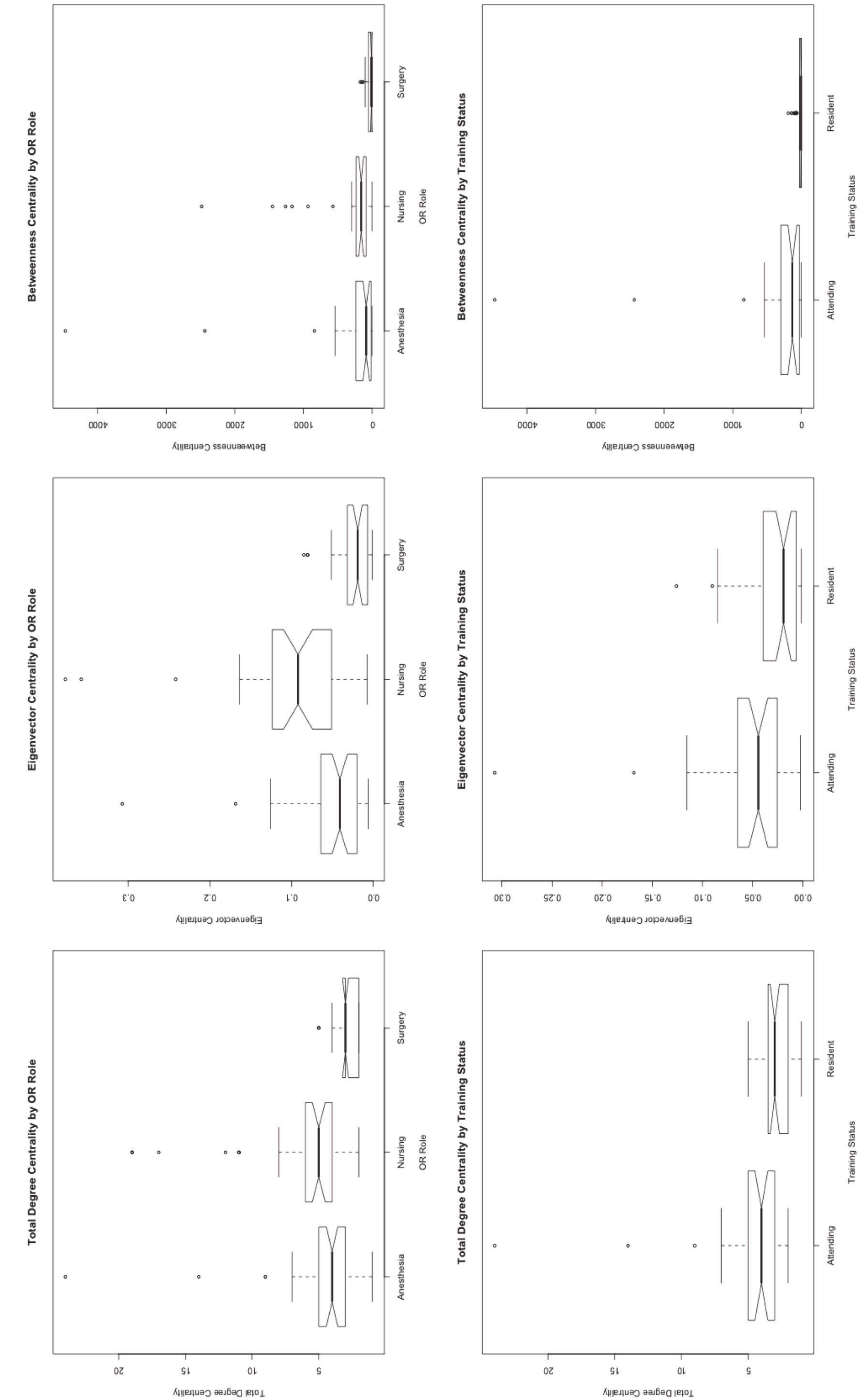


Figure 4: Box Plot of Centrality Scores by Operating Room Role and Physician Training Status

And yet, cardiac surgery is performed by teams of individuals. The focus on anesthesiologist No. 7, while certainly serving as a didactic point on the risks of perioperative tachycardia, fails to appreciate the potential for compensatory interventions by others participating is n No. 7's cases. Such individually-focused approaches to medical error prevention fail to appreciate the individual as a member of a complex healthcare team with potentially offsetting strengths and weaknesses. As network-centric approaches to adverse event prevention gain traction, methods to quantify network complexity and structure at both the micro- and meso-levels will likely become critical (Reason, Carthey, & de Leval, 2001).

6. Limitations

Although this study provides insight into the relative organizational impact of certain roles in the OR, it is based on a single healthcare system and on a single, average day in the OR. This limits the generalizability of our results to other days of the week within our own institution, when other services may have block time, let alone other institutions. This approach also fails to capture information pertaining to the nature of the connections between nodes, especially the rate, volume, and nature of communications among individuals. Nevertheless, this model captures themes common to modern operating rooms, including the hierarchal approach to OR team member connectivity, the parallel roles of surgeons, anesthesiologists and nurses, and relative intraoperative connectivity of attending versus resident physicians. Future work will need to quantify the communiqué-vectors within the network, how much information is transmitted with each communiqué, communiqué duration, and the importance of the included information.

In conclusion, our analysis suggests that an OR at a tertiary-care academic medical center is a scale-free network exhibiting small-world characteristics. Given the homogeneity of certain scale-free network characteristics throughout nature, such a designation may have implications for coordinating anesthesiologists and charge nurses and for understanding the disproportionate impact of OR growth on these roles. In the next phase of this project, we will explore how meso-level organizational network complexity impacts perioperative patient outcomes.

References

Albert, R., & Barabasi, A. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics* 74, January 2002. doi:10.1103/RevModPhys.74.47

Albert, R. Jeong, H., & Barabasi, A. (2000). Error and attack tolerance of complex networks. *Nature*, 406(6794), 378-382. doi:10.1038/35019019

Anderson, C., & Talsma, A. (2011). Characterizing the structure of operating room staffing using social network analysis. *Nursing Research*, 60(6), 378-385. doi:10.1097/NNR.0b013e3182337d97

Anderson, J. G., & Jay, S. J. (1985). The diffusion of medical technology: social network analysis and policy research. *The Sociological Quarterly*, 26(1), 49-64. doi:10.1111/j.1533-8525.1985.tb00215.x

Barabsi, A.L, & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439), 509-512.

Barabási, A. L. (2009). Scale-free networks: a decade and beyond. *Science*, 325(5939), 412-413. doi:10.1126/science.1173299

Baumgart, A., Denz, C., Bender, H., & Schleppers, A. (2009). How work context affects operating room processes: using data mining and computer simulation to analyze facility and process design. *Quality Management in Health Care*, 18(4), 305-314. doi:10.1097/QMH.0b013e3181bee2c6

Bonacich, P. (2007). Some unique properties of eigenvector centrality. *Social Networks*, 29, 555-564.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *The Journal of Mathematical Sociology*, 2(1), 113-120. doi:10.1080/0022250X.1972.9989806

Chambers, D., Wilson, P., Thompson, C., & Harden, M. (2012). Social network analysis in healthcare settings: a systematic scoping review. *PloS one*, 7(8), e41911. doi:10.1371/journal.pone.0041911

Clauset, A, Shalizi, C. R., & Newman, M. (2009). Power-law distributions in empirical data. *SIAM Review* 51, 661-703. DOI: 10.1137/070710111

Cohen, R., Erez, K., be-Avraham, D., & Havlin, S. (2000). Resilience of the Internet to random breakdowns. *Physical Review Letters*, 85(21), 4626-4628. doi:10.1103/PhysRevLett.85.4626

Coleman, J., Katz, E., & Menzel, H. (1957). The diffusion of an innovation among physicians. *Sociometry*, 20, 253-270.

Dalkey, N., & Helmer, O. (1963). An experimental application of the DELPHI method to the use of experts. *Management Science*, 9(3), 458-467. doi:10.1287/mnsc.9.3.458

Effken, J. A., Carley, K. M., Gephart, S., Verran, J. A., Bianchi, D., Reminga, J., & Brewer, B. B. (2011).

Using ORA to explore the relationship of nursing unit communication to patient safety and quality outcomes. *International Journal of Medical Informatics*, 80(7), 507-517. doi:10.1016/j.ijmedinf.2011.03.015

Eichhorn, J. (2013). Practical current issues in perioperative patient safety. *Canadian Journal of Anesthesia*, 60, 111-8. doi:10.1007/s12630-012-9852-z.

ElBardissi, A. W., & Sundt, T. M. (2012). Human factors and operating room safety. *Surgical Clinics of North America*, 92(1), 21-35. doi:10.1016/j.suc.2011.11.007

Erdős, P., & Rényi, A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, 6, 290-297.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215-239. doi:10.1016/0378-8733(78)90021-7

Haynes, A. B., Weiser, T. G., Berry, W. R., Lipsitz, S. R., Breizat, A. S., Dellinger, E. P., et al. (2009). A surgical safety checklist to reduce morbidity and mortality in a global population. *The New England Journal of Medicine*, 360(5), 491-499. doi:10.1056/NEJMsa0810119

Hu, Y. Y., Arriaga, A. F., Peyre, S. E., Corso, K. A., Roth, E. M., Greenberg, C. C. (2012). Deconstructing intraoperative communication failures. *Journal of Surgical Research*, 177, 37-42. doi:10.1016/j.jss.2012.04.029.

Humphries, M. D., & Gurney, K. (2008). Network "small-world-ness": a quantitative method for determining canonical network equivalence *PloS One*, 3(4), e0002051. doi:10.1371/journal.pone.0002051

Klemm, K., & Eguíluz, V. (2002a). Highly clustered scale-free networks. *Phys Rev*, E65, 036123

Klemm, K., & Eguíluz, V. (2002b) Growing scale-free networks with small-world behavior. *Phys Rev*, E65, 057102

Krackhardt, D. (1990). Assessing the political landscape: structure, cognition, and power in organizations. *Administrative Science Quarterly*, 35, 342-369.

Krackhardt, D., & Hanson, J. R. (1993). Informal networks: the company behind the chart. *Harvard Business Review*, 71(4), 104-111.

Krebs, V. E.(2002). Uncloaking terrorist networks. *First Monday 7*, Number 4-1, at http://journals.uic.edu/ojs/index.php/fm/article/view/941/863

Listyowardojo, T. A.(2012). Improving safety culture in health care: implications of individual and institutional variability, at http://dissertations.ub.rug.nl/FILES/faculties/gmw/2012/t.a.listyowardojo/thesis.pdf#page=71

Nakano, T., & White, D. R. (2006). The large-scale network of a Tokyo industrial district: small-world, scale-free, or depth hierarchy *Citeseer, 1100*, 3355.

Reason, J. T., Carthey, J., & de Leval, M. R. (2001). Diagnosing "vulnerable system syndrome": an essential prerequisite to effective risk management Quality in Health Care: QHC, 10 Suppl 2, ii21-25.

Slogoff, S, & Keats, A. S. (1985). Does perioperative myocardial ischemia lead to postoperative myocardial infarction *Anesthesiology*, 62(2), 107-114.

Slogoff, S., & Keats, A. S. (2006). Myocardial ischemia revisited. *Anesthesiology*, 105(1), 214-216.

Small, M., & Chi, K. (2005). Small world and scale free model of transmission of SARS. International Journal of Bifurcation and Chaos, 15(05), 1745-1755. http://www.worldscientific.com/doi/abs/10.1142/S0218127405012776

Stulberg, J. J., Delaney, C. P., Neuhauser, D. V., Aron, D. C., Fu, P., & Koroukian, S. M. (2010). Adherence to surgical care improvement project measures and the association with postoperative infections JAMA: *The Journal of the American Medical Association*, 303(24), 2479-2485. doi:10.1001/jama.2010.841

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of "small-world" networks *Nature*, 393(6684), 440-442. doi:10.1038/30918

Woloshynowych, M., Davis, R., Brown, R., Vincent, C. (2007). Communication patterns in a UK emergency department. *Annals of Emergency Medicine*, 50, 407-413. doi:10.1016/j.annemergmed.2007.08.001.

Xu, J., Hu, D., & Chen, H. (2009). The dynamics of terrorist networks: understanding the survival mechanisms of global Salafi Jihad. *Journal of Homeland Security and Emergency Management*, 6(1). doi:10.2202/1547-7355.1477

Appendix A

The power law distribution is defined as:

$$p(x) = \frac{x^{-\alpha}}{\sum_{n=0}^{\infty}(n + x_{min})^{-\alpha}}$$

such that $x_{\_min}$ defines a lower bound and the scaling parameter α is typically in the range of two to three.

Because power law distributions often pertain only to the tail of a distribution, Kolmogorov-Smirnov testing for goodness of fit was performed for those nodes that, when ranked from greatest to least un-scaled total degree centralities, held a rank greater than that of x_{min}. The x_{min} values were obtained using the methods of Clauset, Shalizi, & Newman (2007) using the R package implementation offered by Dubroca. Further testing of the distribution of total degree centrality against the log, lognormal, and exponential distributions was conducted in order to ascertain with greater certainty the power law distribution of the observed total degree centralities.[1]

Reference

1. Clauset A, Shalizi CR, Newman MEJ: Power-law distributions in empirical data. SIAM Rev 2009;51:661-703

Appendix B

Figure Legends

Figure 1. Plot of the Operating Room Communication Network. The OR network which was modeled is comprised of two hospital buildings connected by an underground tunnel. Each hospital has one floor of operating rooms. One of the hospitals (North Tower) also contains obstetric, MRI, endoscopy, and interventional radiology suites in separate remote locations. For the purposes of this model, perioperative staffing included surgeons, surgical residents, anesthesiologist, anesthesiology residents, midlevel anesthetic providers, and charge nurses within the post-anesthesia care units (PACU), intensive care units (ICU) and operating rooms. All staff are able to work in either hospital, however there is no intra-case movement between the two hospitals. Each facility maintains its own preoperative holding area and PACU. Surgical services and subspecialty surgical ICUs are divided among the 2 hospitals (North and South Towers). The North Tower (NT) consists of Cardiothoracic, Vascular, Otolaryngology, Pediatrics, Neurosurgery, and Pediatric Cardiac cases. The South Tower (ST) consists of General Surgery, Transplant, Urology, and Orthopedic cases. The NT houses the Neurosurgical, Pediatric, and Cardiac Intensive Care Units. The ST houses the Surgical Intensive Care Unit which takes care of Trauma, General Surgery, Orthopedics, Vascular, Otolaryngology, Transplant, and Urology patients.

Figure 2. Example Prototypes of Random and Small-World Networks. Erdos-Renyi (Panel A) network created using the same set of 146 nodes from the observed OR network, configured according to an Erdos-Renyi network structure with a density of 0.05. The small-world network (Panel B) was also created using the 146 nodes from the observed OR network, with the number of neighbors set to 10, probability of neighbor removal 0.05, probability of adding far neighbor 0.01, and power law exponent 0.01.

Figure 3. Proportions of OR Network Healthcare Providers by Role and Training Status. Ninety-one nodes (62%) were physicians, 42 (29%) nodes were nurses, and 13 (9%) nodes were CRNA/AAs. Of the physicians, 42 (46% of physicians, 29% of all nodes) were anesthesiologists and 49 (54% of physicians, 34% of all nodes) were surgeons. Twenty-three (55%) of the anesthesiologists, and 25 (51%) of the surgeons, were trainee physicians. The anesthesia team included 13 CRNAs or AAs representing 24% of the anesthesia workforce.

Figure 4. Box Plot of Centrality Scores by Operating Room Role and Physician Training Status. For Total Degree Centrality there were statistically significant differences between Anesthesia and Nursing ($p=0.005$), Anesthesia and Surgery ($p<0.0001$), and Nursing and Surgery ($p<0.0001$). For Eigenvector Centrality, there were statistically significant differences between Anesthesia and Nursing ($p<0.0001$), Anesthesia and Surgery ($p=0.0002$), and Nursing and Surgery ($p<0.0001$). Betweenness Centrality scores differed between Anesthesia and Surgery ($p=0.0004$), between Nursing and Surgery ($p<0.0001$), but not between Nursing and Anesthesia ($p=0.09$). Differences between Attendings and Resident physicians were statistically significant ($p<0.0001$) for Total Degree Centrality, Eigenvector Centrality, and Betweenness Centrality.

Mobile Phone Assessment in Egocentric Networks: A Pilot Study on Gay Men and Their Peers

W. Scott Comulada
University of California
Los Angeles, California

Abstract
Mobile phone-based data collection encompasses the richness of social network research. Both individual-level and network-level measures can be recorded. For example, health-related behaviors can be reported via mobile assessment. Social interactions can be assessed by phone-log data. Yet the potential of mobile phone data collection has largely been untapped. This is especially true of egocentric studies in public health settings where mobile phones can enhance both data collection and intervention delivery, e.g. mobile users can video chat with counselors. This is due in part to privacy issues and other barriers that are more difficult to address outside of academic settings where most mobile research to date has taken place. In this article, we aim to inform a broader discussion on mobile research. In particular, benefits and challenges to mobile phone-based data collection are highlighted through our mobile phone-based pilot study that was conducted on egocentric networks of 12 gay men (n = 44 total participants). HIV-transmission and general health behaviors were reported through a mobile phone-based daily assessment that was administered through study participants' own mobile phones. Phone log information was collected from gay men with Android phones. Benefits and challenges to mobile implementation are discussed, along with the application of multi-level models to the type of longitudinal egocentric data that we collected.

Keywords: Gay men, HIV risk behaviors, mobile phone log, ecological momentary assessment, ohmage

Authors
W. Scott Comulada, is an assistant professor-in-residence in the Department of Psychiatry and Biobehavioral Sciences at the University of California, Los Angeles, California. He is also a Methods Core scientist for the Center for HIV Identification, Prevention, and Treatment Services (CHIPTS; P30MH058107)

Acknowledgements
This study was funded and supported by the National Institute of Mental Health (K01MH089270; P30MH058107). Mobile phone-based data collection was supported by ohmage and the following centers at the University of California, Los Angeles: the Center for Embedded Networked Sensing and Mobilize Labs. In particular, I would like to thank Deborah Estrin and Nithya Ramanathan for input on the development of the mobile assessment.

Please send all correspondence to W. Scott Comulada, UCLA Center for Community Health, 10920 Wilshire Blvd, Suite 350, Los Angeles, CA 90024, USA. wcomulada@mednet.ucla.edu.

1. Introduction

Mobile phones are by nature social devices as highlighted by numerous studies on the structure of mobile communication networks (e.g., Onnela et al., 2007; Ye et al., 2008) and individual tie strengths (Zhang & Dantu, 2010). Most studies have analyzed sociometric data where large and bounded networks are observed. In contrast, egocentric networks of individuals, i.e. *egos*, and their peers, i.e. *alters*, have received less attention. In part, the focus on sociometric data may be due to the availability of "safe" data sets that are collected in laboratory settings, often on faculty and students where privacy issues are less critical and participants are technologically savvy. A good example is the MIT Reality Mining data set (Zhang & Dantu, 2010).

Egocentric networks are often assessed in public health settings on marginalized populations, e.g. drug-using networks (Yang, Latkin, Muth, & Rudolph, 2013). Privacy is critical and "tech-savvy" assumptions may be unrealistic. Yet mobile assessment has been successfully carried out in cocaine-addicted homeless patients (Freedman et al., 2006) and other marginalized populations. Furthermore, mobile technologies can enhance data collection and intervention delivery in public health settings, e.g. ecological momentary interventions (Heron & Smyth, 2010). As we have found in our own transition from more traditional modes of data collection to mobile-based studies, an important part of the implementation process is a clear understanding of what mobile technologies can and cannot do. As noted by Lazer et al. (2009), researchers and institutional review boards (IRBs) alike need to be up to speed on the latest technologies in order to design and evaluate proper privacy and encryption protocols, respectively.

In this article, we highlight the benefits and limitations of mobile data collection through egocentric data that was collected to test the implementation of a mobile phone-based health assessment in a sample of 12 gay men, i.e. *egos*, and their peers, i.e. *alters*. Both egos and alters used their own phones to fill out a health assessment and enter sensitive information on HIV-transmission behaviors. We collected phone-log data from a subset of the egos with Android phones in order to compare mobile communications with alters in the study and with individuals who did not enroll in the study. Therefore, our study provides a good opportunity to discuss privacy and ethical issues that are central to public health settings.

We also give examples of research questions and analytic strategies that are afforded by the collection of mobile data in an egocentric study. A key feature of our data is the three levels of hierarchy. Egocentric data normally contains two levels where individuals (both egos and alters) are nested within egocentric networks. Multi-level models are applied and contain random effects for each network to allow mean levels of the outcome to differ across networks (e.g., Hall, 2010; Rice et al., 2009; Snijders, Spreen, & Zwaagstra, 1995; Valente, 2010). In our study, participants filled out an end-of-the-day mobile assessment over a month; repeated observations are nested within individuals. We discuss extensions to the basic multi-level model to analyze longitudinal egocentric data. It is important to note that longitudinal data in our study resulted from daily reporting which is a course version of ecological momentary assessment (EMA) where events are recorded as they occur in situ. EMA also involves a large number of repeated measurements and depends on careful timing, e.g. several times a day, to capture variations in behavior within days (See Shiffman, Stone, & Hufford, 2008, and Stone and Shiffman, 1994, for overviews). In contrast, standard assessment methods rely on retrospective recall where study participants are asked to report on behaviors over a period of time and are often interviewed in a clinical setting. EMA minimizes recall biases that are intensified as individuals reconstruct and retrieve events from their memory over longer periods of time. By self-administering assessments, EMA may also reduce interview bias, e.g. in giving socially desirable responses to sexual behavior questions (Kissinger et al., 1999).

2. Data and Methods

2.1 Participants

Recruitment was conducted online (Figure 1). From April to August, 2013, 455 egos were recruited through pop-up messages on *Grindr*, a dating website for gay men, and postings on *Craigslist* that directed them to a study webpage. *Craigslist* is an online forum for classified ads.

The study webpage directed egos to online screening and consent forms that were hosted by *SurveyMonkey* (http://www.surveymonkey.com/). Study eligibility required egos to 1) self-identify as a gay or bisexual man; 2) be at least 18 years old; 3) live in Los Angeles County; 4) use a web-enabled Android phone, version 2.3 or higher (issued after November 2010), or an iPhone; 5) use their mobile phone to participate in the study; and 6) recruit at least 3 alters who had an Android or iPhone they could use to participate in the study.

Out of 455 egos who started the online forms, 19% were not eligible (n = 85) and 37% did not finish filling out the forms (n = 167). It is hard to know why

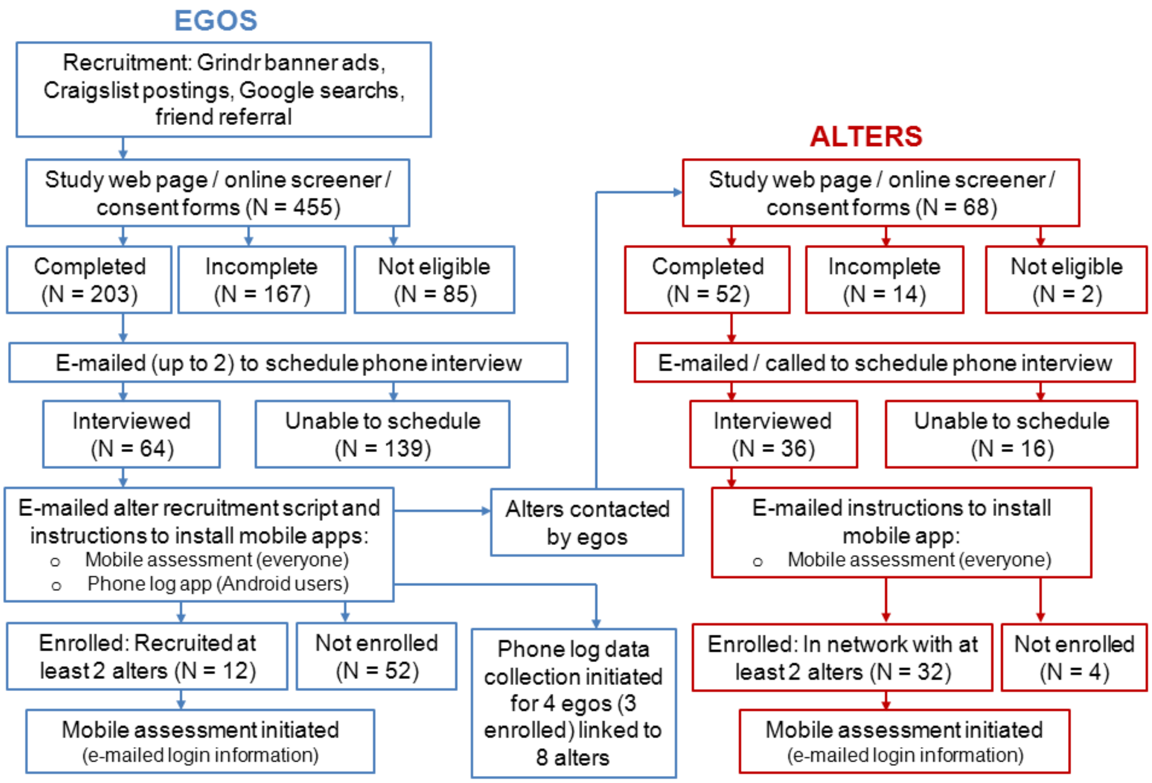


Figure 1: Recruitment and enrollment of egos (gay men) and alters and initiation of mobile phone-based data collection.

so many individuals did not finish filling out forms. In at least one instance, a peer started to fill out the forms on their mobile phone, lost internet connection, and did not attempt to re-initiate the forms.

Eligible egos (45%; n = 203 of 455) were e-mailed to set up a one-time telephone interview and also received instructions on how to install and use the study mobile apps; calls were scheduled with 64 egos. During the call, we administered a demographic and social network assessment. *Grindr* banner ads were the primary recruitment source (n = 53 of 64 telephone interviews).

After the telephone interview, egos were sent an e-mail template they could send to alters they wished to invite into the study. The e-mail template contained a link that directed interested alters to a separate study webpage and in turn, online screener and consent forms. The online form asked alters to enter the first name and phone number of the ego who recruited them so we could construct ego-alter links. Eligible alters fulfilling 2), 4), and 5) were contacted and administered a demographic assessment. We relaxed the requirement for egos to recruit 3 alters and allowed egos and alters to participate if at least 2 alters per egocentric network were recruited. Out of 64 egos who completed a telephone interview, roughly 1 in 5 recruited at least 2 alters and enrolled in the study (n = 12 of 64; Figure 1). We did not follow-up with unenrolled egos to find out the reason. One ego let us know that his friends did not want to join and "share private information".

Out of 68 alters who started the online screener, 75% (n=51) completed the screener and provided contact information to schedule an interview. Thirty two of 36 alters who were interviewed by telephone enrolled in the study.

Egos and alters were e-mailed Amazon gift card activation codes worth \$60 and \$50, respectively, at the end of the study as incentives. Egos and alters who were the most compliant in filling out the daily health assessment were entered into a drawing to also receive an Amazon gift card activation code worth \$100. All study procedures were approved by the Institutional Review Board at the University of California, Los Angeles.

2.2 Data collection

Telephone interviews were conducted at the beginning of the study prior to the start of the mobile phone health assessment. Egos were queried on where they heard about the study, the model of the mobile phone they would be using during the study, age, and ethnicity. Alters were queried on their relationship to the ego who recruited them, gender, age, ethnicity, whether or not they lived in Los Angeles County, and their sexual orientation. During the telephone interview, egos were also administered a 9-item adapted version of the Arizona Social Support Inventory (Barrera & Gottlieb, 1981) to elicit names of people with whom the respondent socializes, lives, eats meals, has sex, does alcohol and drugs, receives health advice, calls upon for material and emotional support, or any other people who were important to them that had not been prompted by the prior name-generator questions. After the 9-item inventory, we asked for names of alters the ego was planning to ask to join them in the study. Almost all of the egos who enrolled in the study recruited at least one alter who had not been prompted by the 9-item inventory (n = 10 of 12). We calculated the size of each egocentric network based on the number of names generated by the 9-item inventory. Seven of the 64 egos gave "other" responses that encompassed multiple people, e.g. "family", and were excluded from the network-size calculation.

Phone logs were recorded through *SystemSens* (http://systemsens.ohmage.org), an Android application that was designed to collect passive system data and developed through the UCLA Center for Embedded Networked Sensing. Egos with Android phones were asked to download *SystemSens* to their mobile phone through an e-mail link. Once installed, *SystemSens* automatically encrypted and uploaded phone-log data (including phone numbers, the duration, and date / time stamp of incoming and outgoing calls and text messages) to servers at UCLA whenever the user charged their phone. To protect the identity of phone numbers belonging to individuals who were not enrolled in the study, all phone numbers in the phone log were scrambled using SHA-256, a cryptographic hash function published by the National Institute of Standards and Technology. There are several notable features of SHA-256. Hashed numbers appear as unique 256-bit values, e.g. "b8475260a8bdd4af2984d7d7d8eb9b5a". As a result, one is able to identify if two hashed numbers are of the same phone number. However, it is nearly impossible to recover the original phone number from a hashed number alone. The original phone number is necessary to act as a key that unscrambles the hashed number and verifies the original number.

Mobile phone assessment. All participants (both egos and alters) were asked to fill out the same daily assessment on their mobile phone for a month. Assessments were launched using the *ohmage* application (http://www.ohmage.org), an open-source application that is compatible with Android and iPhones. *Ohmage* allows for assessments to be rapidly authored using the Extensible Markup Language (XML), and allows data to flow from participants' mobile phones to a centralized database. In this study, *ohmage* was launched with an HTML5 application implemented using the Mobile Web Framework (MWF). The application runs on both Android and iPhones and is available for download from the Google Play and Apple app stores, respectively. A version of *ohmage* that is native to Android phones has been implemented in prior studies (Swendeman et al., 2014); feedback from focus groups on prior mobile studies informed the design of the mobile assessment (Ramanathan et al., 2012). Once installed, participants accessed the mobile asssessment through the *ohmage* dashboard shown in Figure 2A. At the end of each assessment, responses were encrypted, uploaded to servers at UCLA, and removed from the user's mobile phone, as long as there was network connectivity and the phone battery was not low. Responses could also be manually uploaded at a later time.

Mobile phone assessment consisted of 14 questions that participants were asked to fill out at the end of the day for a month. Questions encompassed the

Table 1: Frequency of communication with alters based on ego reports (includes face-to-face, telephone, and social media contact) and based on the number of days between phone log calls/ text messages between egos and alters

				Call logs	
Ego	Alter	Alter relation	Self-report	Median days	Range
1	1	Partner	Daily	1	1-3
	2	Friend	Once a week	2.5	1-13
	3	Friend	Once a week	7	2-9
2	1	Boyfriend	Daily	1	1-5
	2	Friend	Daily	1	1-2
3	1	Partner	Daily	8.5	7-10
4	1	Ex-boyfriend	3.5 times a week	1	1-7
	2	Sister	Daily	1	1-5

following domains in the following order: (a) An adapted version of the Healthy Days Symptoms Module from the Health-Related Quality of Life instrument (HRQOL; Centers for Disease Control and Prevention, 1995), including 5 questions on mood, worry, sleep, energy level, and impairment; (b) Daily minutes of exercise and type of exercise, e.g. "Jogging"; (c) Rating of one's daily eating, e.g. "Less healthy than usual"; (d) A food inventory that was constructed from multiple food inventories (e.g., Fulkerson et al., 2008; Kaiser et al., 2003; Sisk, Sharkey, McIntosh, & Anding, 2010) and designed to fit across two cell phone screens; (e) Sexual behavior, including the number of sexual encounters involving anal or vaginal sex, the number of encounters with "casual (including one-time and first-time) partners", and condom usage; and (g) Alcohol and substance use.

All questions included a "Refuse to answer" response option so that participants were not forced to answer any questions they did not want to. However, we did not want participants to repetitively select refusal responses in order to get through the daily assessment more quickly. We placed additional "speed bump" questions that required participants to specify why they refused to answer the prior question in two places. The first speed-bump question was placed after minutes of exercise were queried, and the second was placed after the number of sexual encounters was queried at approximately the halfway point and end of the assessment. No refusals were entered, except for the impairment question (1 refusal) and substance use (3 refusals).

3. Analytic Strategies and Results

3.1. Sample characteristics

Among egos who were interviewed over the telephone (n = 64), the average age was 30.8 years old (range = 18 to 58). Ethnicity was reported as African American (12.5%), Latino (34.4%), White (37.5%), or Other (15.6%). Egos reported a network size of 8.4 members, on average (range = 2 to 30). Networks were fairly homogenous with respect to age and ethnicity. For example, most of the White egos (n = 5 of 6) only recruited White alters. Half of the Latino egos (n = 2 of 4) only recruited Latino alters.

3.2. Call logs

Phone logs were recorded for four egos with Android phones. Logs began recording as soon as SystemSens was installed and continued until the end of the study that included the 30-day health assessment time period

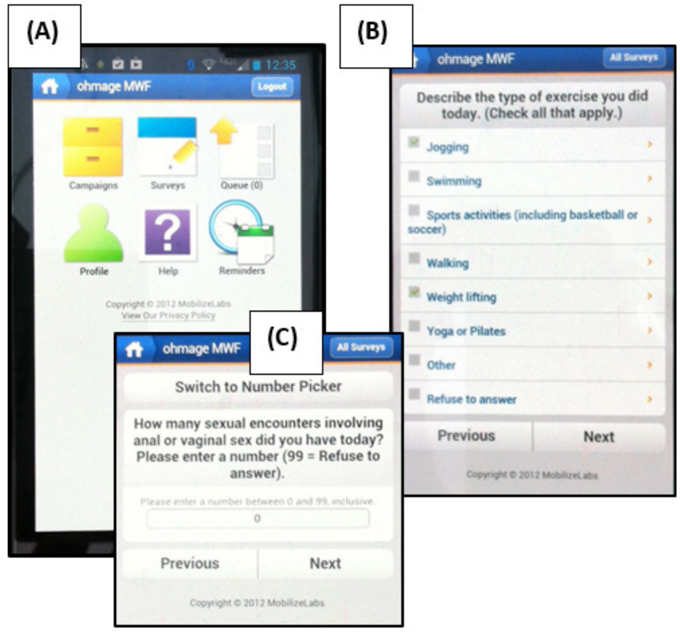


Figure 2: ohmage MWF screenshots showing (A) dashboard for accessing daily health survey and sample questions from the daily health survey, including a (B) multiple-response item and (C) an item requiring numeric entry.

(range = 20 to 45 days). One ego was only able to recruit one peer and dropped out of the study after 20 days. Similar to Onnela et al. (2007), we excluded one-way communications where calls or text from an ego to a phone number occurred, or vice versa, but were not reciprocated. By focusing on reciprocated communications, we eliminated communications related to single events where egos did not personally know individuals they were communicating with. Two phone log analyses are discussed.

Agreement between self-reported contact and phone logs. The frequency of contact with network members is typically self-reported by egos. Given the additional contact information provided by mobile communication (both calls and text messages), a natural question arises. Do phone logs provide overlapping information to self-reported contact or do phone logs provide additional information? Table 1 demonstrates a way to address this question by showing egos' self-reported frequency of contact with alters that was reported during the telephone interview and the median number of days between mobile communications with alters. Phone logs corroborate the self-reported frequencies fairly well. For example, four of five "daily" reports matched up with call logs where half of the communications occurred within a day of each other.

Alter closeness. There is a general understanding in social network research that observed networks in a study are incomplete. Social ties with individuals outside

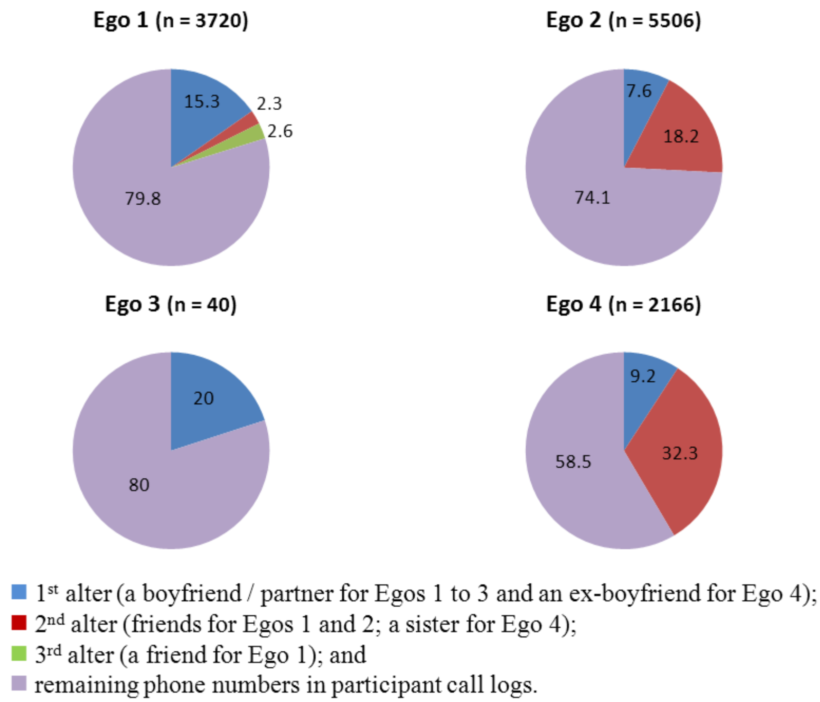


Figure 3: Percentage of (n) reciprocated calls or text messages:

the study network can sometimes be constructed by self-report (e.g., Fowler & Christakis, 2008), though this is typically not the case. Therefore, phone log communication data can fill in gaps on self-reported network compositions. In particular, we focus on the frequency of egos' mobile communications with recruited alters and individuals outside the study as a proxy for ego-alter closeness. Information on closeness with alters who are likely to be recruited into a study has the potential to inform the design of both social network-based interventions (see Valente, 2012 for a review) and recruitment strategies (e.g., respondent-driven sampling; Heckathorn, 1997, 2002). Figure 3 shows the percentage of communications with each alter and with the remaining telephone numbers in the phone logs. Among these four egos, we note that they recruited at least one alter they were in fairly frequent contact with, e.g. partners for Egos 1 and 3 (15.3% and 20.0% of the total communications, respectively).

3.3. Mobile health assessment

We discuss two types of multi-level regression models that address research questions specific to each level of hierarchy in a longitudinal egocentric data set.

Network-level questions. Holistic health approaches often track multiple and disparate measures of health. For example, le Roux et al. (2013) examined mental health, general health, and HIV-transmission behaviors. In this vein, we examined how multiple health behaviors and HRQOL cluster within networks. Due to the small sample size, an ad hoc approach was used. Responses for each individual were aggregated over their 30-day study period. We then fit separate multi-level models to each HRQOL or behavioral measure. Pearson product-moment correlations were then examined within the 12 pairs of network-level random effects between all possible pairs of HRQOL and behavioral outcomes. Correlations were in expected directions. For example, at the network level, there were negative correlations between numbers of alcoholic beverages and both mean levels of healthy feelings ($r = -.42$) and days of exercise ($r = -.50$).

A more formal modeling approach uses a bivariate-outcome multi-level model similar to Comulada et al. (2010, 2012). Here we consider longitudinal egocentric data with two continuous outcome measures, e.g. levels of mood and sleep. For individual i in network n at time point t and outcome k (= 1,2), a bivariate random-intercept linear model on continuous outcome y_{nitk} is expressed as

$$y_{nitk} = x_{nitk}'\beta_k + \lambda_{nk} + \eta_{nik} + \varepsilon_{nitk}, \quad (1)$$

where β_k is a vector of regression coefficients for covariate vector x_{nitk} on outcome k. Correlations for each outcome within networks and across repeated observations within individuals are accounted for by random effects λ_{nk} and η_{nik}, respectively. Residual error term ε_{nitk} accounts for variance that is unexplained by the random effects. A key feature of the model is that correlations between outcomes are modeled through a variance-covariance matrix that is shared by random effects and residual terms across outcomes. In particular, cross-correlations can be examined between outcomes at different time points, e.g. the relationship between drug use and trust between egos and alters over several time points (Comulada et al., 2012).

Individual-level questions. Longitudinal studies typically entail a few time points. Analyses focus on mean changes over time, e.g. decreases in drug use. EMA in our study resulted in numerous time points (intensive longitudinal data; Walls & Shafer, 2006). In larger samples, changes in variability, as well as mean levels, can be examined using location scale models (Hedeker, Mermelstein, & Demirtas, 2008; Hedeker, Demirtas, & Mermelstein, 2009). For example, Hedeker, Demirtas, & Mermelstein (2009) examined mood fluctuations in smokers over time.

4. Discussion

Our mobile phone-based pilot study on egocentric networks of gay men and their peers highlights a number of benefits that are scalable to larger studies and other populations.

First, recruitment and implementation of the study was carried out without in-person visits with study participants. Second, participants used their own mobile phones, which alleviated the need to carry another electronic data-entry device. Both features served to reduce participant burden and study costs that are associated with traditional studies, e.g. interviewers were not needed. A degree of anonymity was also provided for participants, which may be an important issue for marginalized populations.

Past EMA studies have typically relied on paper diaries that are prone to backfilling (Stone et al., 2002, 2003). Palm top computers address this issue, but still introduce a degree of user burden that can be attenuated by making use of an individual's own mobile phone. An important feature of our study, in terms of data quality, was the ability to visualize uploaded mobile assessment data through a website portal in near real time. Research staff checked the data every few days. In one instance, no data was observed for an alter after initial study enrollment. Telephone contact with the study participant revealed that they were accidentally preventing their assessment data from being uploaded to the study team. The problem was easily corrected, and only a few days of data were lost.

The strength of our technologically-driven study design is also an obvious limitation for implementation in other populations. In studying egocentric networks of gay men who use *Grindr* and live in Los Angeles, we focused on a fairly tech-savvy population. Furthermore, gay men in Los Angeles are often targeted for HIV-related studies, especially through *Grindr* (e.g., Rendina et al., 2014). At enrollment, a number of our study participants were already familiar with standard study protocols. These characteristics facilitated the use of online recruitment and mobile assessment. Using these tools would be more difficult in other populations where study details are better explained in person and where study participants may be more reluctant to enter sensitive information during a survey, especially on an electronic device. Few concerns were voiced by participants in our study. Online recruitment may be unethical in populations were a language barrier is present, and online consent forms may be easy to click through without understanding the content.

Despite the technological savvy of our population, three main limitations remained with our study design. Approximately half of the eligible gay men who clicked through our *Grindr* banner ad and initiated the online forms, completed the study participation forms (55%; n = 203 / 370; Figure 1). This percentage is similar to initial participation rates that were found in another study that recruited gay men through *Grindr* and asked them to fill out a one-time online survey (43%; n = 2175 / 5026; Rendina et al., 2014). A big difference between Rendina et al. (2014) and our study is that they retained 27% of the initial gay men in their analysis sample. We retained 12 egos (3%) in our study. Increasing rates of online recruitment offers potential participants a smorgasbord of studies to select from. Moreover, there is less buy-in when shopping amongst online studies.

For example, rapport may be established with a recruiter during recruitment in a clinic. Online recruitment may be best suited for studies offering instant participation. Recruitment through Grindr reached gay men with risky sexual behavior profiles as intended; Nineteen percent of interviewed gay men reported anonymous / one-time sex partners in their network during the telephone interview (n = 12 of 64). Yet only one of the 12 (8%) enrolled gay men had reported anonymous partners. Though not statistically significant, this percentage drop suggests that an online forum that attracts users with a targeted behavior is not necessarily a good recruitment source.

Another limitation was our restriction of phone-log data collection to Android users. In our study, the majority of participants were iPhone users, e.g. 64% of egos and 70% of alters who filled out online forms. iPhone users tend to have other iPhone users as friends (Canright, 2013). Android and iPhone users also tend to have different demographic and social characteristics (Albanesius, 2011). Phone log-based inference that is based on one type of mobile phone is likely to miss a segment of the population and be biased. Text message-based assessment that that does not require a smartphone may be a better option in other populations.

Lastly, lengthy assessments may call for larger computer screens and human interaction to encourage compliance. Our mobile assessment could be taken in a few minutes. Questions contained a few response categories and mostly fit on one screen. This may partly explain high compliance in our study (a median of 24 days of reporting).

The benefits and challenges in our study support a marriage of traditional and new data collection methods that is likely to remain in social network research. Visual web interfaces that allow participants to construct their own personal networks through a self-administered social network inventory have met with limited success; an interviewer may still be necessary (Matzat & Snijders, 2013). Moreover, one mode of electronic communication may not adequately capture social interaction (Quintane & Kleinbaum, 2011). That is why we assessed the frequency of ego-alter contacts through self-report and mobile communication. Despite the challenges of incorporating new technologies into research, the social dynamics of mobile devices and social media are difficult to ignore. It is hard to fully understand the dynamics of health-related behaviors without them.

References

Albanesius, C. (2011). Infographic: What does the average Android, iOS user look like? pcmag.com. Accessed October 22, 2013. http://www.pcmag.com/article2/0,2817,2391339,00.asp.

Barrera Jr., M., & Gottlieb, B.H. (1981). Social support in the adjustment of pregnant adolescents: assessment issues. In: Gottlieb, B.H. (Ed.), Social Networks and Social Support. Sage Publications, Beverly Hills, CA, pp. 69–96.

Canright, G. (2013). Who cares what kind of smartphone their friends have? Conference presentation, International Network for Social Network Analysts, SunBelt XXXIII, May 21-26, Hamburg, Germany.

Centers for Disease Control and Prevention. (1995). http://www.cdc.gov/hrqol/hrqol14_measure.htm.

Comulada, W.S., Rotheram-Borus, M.J., Pequegnat, W., Weiss, R.E., Desmond, K.A., Arnold, E.M., Remien, R.H., Morin, S.F., Weinhardt, L.S., Johnson, M.O., & Chesney, M.A. (2010). Relationships over time between mental health symptoms and transmission risk among persons living with HIV. *Psychology of Addictive Behaviors*, 24, 109-18. doi: 10.1037/a0018190.

Comulada, W.S., Muth, S.Q., & Latkin, C.A. (2012). The analysis of multiple ties in longitudinal egocentric network data: A case study on bidirectional relationships between trust and drug use. *Social Networks*, 34, 691-700.

Fowler, J.H., & Christakis, N.A. (2008). Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal*, 337, a2338. doi: 10.1136/bmj.a2338.

Freedman, M.J., Lester, K.M., McNamara, C., Milby, J.B., & Schumacher, J.E. (2006). Cell phones for ecological momentary assessment with cocaine-addicted homeless patients in treatment. *Journal of Substance Abuse Treatment*, 30, 105-111.

Fulkerson, J.A., Nelson, M.C., Lytle, L.A., Moe, S., Heitzler, C., & Pasch, K.E. (2008). The validation of a home food inventory. *International Journal of Behavioral Nutrition and Physical Activity*, 5, 55.

Hall, Jeffrey, A. (2010). Parents' networks: Egocentric networks and unique and shared sources of social support. *Connections*, 30, 41-49.

Heckathorn, D.D. (1997). Respondent driven sampling: A new approach to the study of hidden populations. *Social Problems*, 44, 174-199.

Heckathorn, D.D. (2002). Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems*, 49, 11-34.

Hedeker, D., Mermelstein, R.J., & Demirtas, H. (2008). An application of a mixed-effects location scale model for analysis of ecological momentary assessment (EMA) data. *Biometrics*, 64, 627-634.

Hedeker, D., Demirtas, H., & Mermelstein, R. J. (2009). A mixed ordinal location scale model for analysis of Ecological Momentary Assessment (EMA) data. *Statistics and Its Interface*, 2, 391-401.

Heron, K.E., & Smyth, J.M. (2010). Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. *Br J Health Psychol*, 15(Pt 1), 1–39. doi.org/10.1348/135910709X466063.

Kaiser, L.L., Megar-Quiñonez, H., Townsend, M.S., Nicholson, Y., Fujii, M.L., Martin, A.C., & Lamp, C.L. (2003). Food insecurity and food supplies in Latino households with young children. *Journal of Nutrition Education and Behavior*, 35, 148-153.

Kissinger, P., Rice, J., Farley, T., Trim, S., Jewitt, K., Margavio, V., & Martin, D.H. (1999). Application of computer-assisted interviews to sexual behavior research. *American Journal of Epidemiology*, 149, 950-954.

Lazer, D., Pentland, A(S)., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., & Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. *Science*, 323(5915), 721-723. doi: 10.1126/science.1167742.

le Roux, I.M., Tomlinson, M., Harwood, J.M., O'Connor, M.J., Worthman, C.M., Mbewu, N., Stewart, J., Hartley, M., Swendeman, D., Comulada, W.S., Weiss, R.E., & Rotheram-Borus, M.J. (2013). Outcomes of home visits for pregnant mothers and their infants: a cluster randomized controlled trial. *AIDS*, 27, 1461–1471.

Matzat, U., & Snijders, C. (2013). The quality of ego-centered social network data in web surveys: experiments with a visual elicitation method. Conference presentation, International Network for Social Network Analysts, SunBelt XXXIII, May 21-26, Hamburg, Germany.

Onnela, J.P, Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., & Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18), 7332-7336.

Quintane, E., & Kleinbaum, A.M. (2011). Matter over Mind? E-mail data and the measurement of social networks. *Connections*, 31(1), 20-41.

Ramanathan, N., Swendeman, D., Comulada, W.S., Estrin, D., & Rotheram-Borus, M.J. (2012). Identifying preferences for mobile health applications for self-monitoring and self-management: Focus group findings from HIV-positive persons and young mothers. *International Journal of Medical Informatics*, 82, e38-46. doi: 10.1016/j.ijmedinf.2012.05.009.

Rice, E., Comulada, S., Green, S., Arnold, E.M., & Rotheram-Borus, M.J. (2009). Differential disclosure across social network ties among women living with HIV. *AIDS and Behavior*, 13, 1253-1261.

Shiffman, S., Stone, A.A., & Hufford, M.R. (2008). Ecological momentary assessment. *Annu Rev Clin Psychol*, 4, 1-32.

Sisk, C., Sharkey, J.R., McIntosh, W.A., & Anding, J. (2010). Using multiple household food inventories to measure food availability in the home over 30 days: a pilot study. *Nutrition Journal*, 9, 19. doi:10.1186/1475-2891-9-19.

Snijders, T., Spreen, M., & Zwaagstra, R. (1995). The use of multilevel modeling for analysing personal networks: Networks of cocaine users in an urban area. *Journal of Quantitative Anthropology*, 5, 85-105.

Stone, A.A., & Shiffman, S. (1994). Ecological momentary assessment in behavioral medicine. *Annals of Behavioral Medicine*, 16, 199-202.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2002). Patient non-compliance with paper diaries. *BMJ*, 324, 1193-4.

Stone, A. A., Shiffman, S., Schwartz, J. E., Broderick, J. E., & Hufford, M. R. (2003). Patient compliance with paper and electronic diaries. *Controlled Clinical Trials*, 24, 182-99.

Swendeman, D., Comulada, W.S., Ramanathan, N., Lazar, M, & Estrin, D. Reliability and validity of daily self-monitoring by smartphone application for health-related quality of life, antiretroviral adherence, substance use, and sexual behaviors among people living with HIV. . (2014). *AIDS & Behavior*, Epub ahead of print.

Valente, T.W. (2010). Social Networks and Health: Models, Methods, and Applications. New York: Oxford University Press. pp. 70-72.

Valente, T.W. (2012). Network interventions. *Science*, 337, 49-53.

Walls, T.A., & Shafer, J.L. (2006). Models for Intensive Longitudinal Data. Oxford University Press.

Yang, C., Latkin, C., Muth, S.Q., & Rudolph, A. (2013). Injection drug users' involvement in drug economy: dynamics of sociometric and egocentric social networks. *Connections*, 33, 24-34.

Ye, Q., Zhu, T., Hu, D., Wu, B., Du, N., & Wang, B. (2008). Cell phone mini challenge award: Social network accuracy – exploring temporal communication in mobile call graphs. IEEE Symposium on Visual Analytics Science and Technology, October 19-24, Columbus, Ohio, USA. doi: 10.1109/VAST.2008.4677389.

Zhang, H., & Dantu, R. (2010). Predicting social ties in mobile phone networks. Intelligence and Security Informatics (ISI), May 23-26, Vancouver, BC, Canada. doi: 10.1109/ISI.2010.5484780.

DEN
Data Exchange Network

The 2012 Malian Conflict Network

Olivier Walther
University of Southern Denmark
Denmark and Rutgers University,
USA

Dimitris Christopoulos
MODUL University Vienna &
CEPS
Luxembourg

Please send all correspondence to ow@sam.sdu.dk.

1. Overview

The 2012 Malian Conflict Network is a dataset collected to illuminate the relationships between individuals affiliated with terrorists and rebel groups in the north of Mali in West Africa. The main objective of the data collection was to study to what extent the terrorists and the rebels were connected by strong brokers and how these two subgroups responded to the recent death of several prominent leaders. Based on open source data from the media, the Malian Conflict Network offers the opportunity to analyze a group of actors, such as Al-Qaeda in the Islamic Maghreb (AQIM) affiliated individuals, who normally operates through a covert network. Temporarily allied to the Tuareg-dominated National Movement for the Liberation of Azawad (MNLA), these terrorists have found in the north of Mali a relatively safe haven for developing their criminal and ideological activities.

2. Data collection

Our network analysis is based on the analysis of a selection of articles published between July 2010 and September 2012 by the French daily Le Monde (27 items) and the weekly Jeune Afrique (30 articles). The period between 2010 and 2012 covers not only the recent conflict between the Malian army, the Tuareg rebellion and the AQMI affiliated terrorists, but also some of the earlier violent events that include killings and abductions in the Sahel-Sahara region. This selection of articles was supplemented by a review of 25 articles published on African news websites such as African 1, Tamtaminfo, Sahara Media, Occitan Touareg and Maliactu, and El Watan.

We then looked for all the names and surnames contained in this corpus of 82 texts. Because our analysis focuses on the relationships between Islamists and rebels we have deliberately ignored the names of government politicians, soldiers from the Malian army and representatives of regional organizations. Our database is composed of 42 actors including 27 Islamists and 15 rebels.

Finally, we identified whether any of the 42 actors were connected to any of the others. We determined that a tie existed between two actors if they had participated in a common political or military event, whatever the duration or location of the encounter. Such operational ties result, for example, from a political meeting, a training in Afghanistan, Iraq or Libya, a participation in combat, a negotiation for hostage release, or an involvement with a killing, an abduction or a bombing.

3. Data files and formats

The data is provided in one Excel Workbook, called MalianNetwork.xlsx, containing two worksheets.

The first worksheet contains the relational data in the form of a matrix of positive and negative ties between social actors. If two actors are mentioned together in the same newspaper article we assume an alliance at degree

1. If there is a mention of a specific strong alliance then the tie has value 2. If there is a mention of two belonging to different factions/religious/ethnic groups we assume a value of -1 and if they are mentioned as enemies a value of -2.

The second worksheet contains attributed data. The following attributes are provided, forming a 42x14 matrix:

- Code
- Name
- Subgroup. Actors are divided between 'Terrorists' and 'Rebels'.
- Organization. The dataset identifies 7 different organizations: Al-Qaeda in the Islamic Maghreb ('AQIM'), the Movement for Oneness and Jihad in West Africa ('Mujao'), 'Ansar al-Dine', the Salafist Group for Preaching and Combat ('GSPC') the Libyan Islamic Fighting Group ('LIFG'), the National Movement for the Liberation of Azawad ('MNLA'), and the May 23, 2006 Democratic Alliance for Change ('ADC').
- Status. Since our analysis is restricted to the last two years of the Malian conflict, we did not consider individuals who deceased, were captured or surrendered before 2010. The data distinguishes between 'Alive', 'Deceased', 'Captured', and 'Unknown'.
- Role. This attribute describes the political or military role of each individual within his organization.
- Nationality

Alias1, 2, 3, 4, 5, 6, 7. Because the spelling of names can greatly vary and because terrorists are well known for using noms de guerre, a list of potential aliases is provided for each, coded from Alias1 to Alias7.

4. Data Details	
Response Rate	N/A
Non-Respondent Bias	N/A
Theoretical Grouping	N/A
Publication Using These Data	Walther O, Christopoulos D. 2014. Islamic Terrorism and the Malian Rebellion. Terrorism and Political Violence 26(2).
Data Context	Territorial conflict in West Africa
Respondents	Members of terrorist groups and of the Tuareg rebellion
Longitudinal	From July 2010 to September 2012
Temporality	Ties between social actors are highly volatile due to the fragile political and ideological alliances. A military intervention initiated by West African states, with the support of their Western allies, has strongly affected the ties observed in the region.
Analytical or Pedagogical Utility	• Demonstrating how terrorists and rebels are interconnected • Showing how terrorists and rebels are internally structured • Illuminating prominent brokers • Illustrating triadic relations • Balance analysis
Known Issues	Data is based on ties reported in newspapers and do not depict actual relations. As is often the case in analyzing terrorist networks from publicly available sources, missing actors or missing links are likely to alter the centrality measures conducted in our work, which calls for caution regarding the interpretation of the rank of each actor.

The "Barras Bravas" Dataset

Javier S. Bundio
University of Buenos Aires
Buenos Aires, Argentina

Please send all correspondence to jsbundio@sociales.uba.ar.

1. Overview

The *barras bravas* (Argentine hooligans) are the most fanatical and violent soccer supporter groups in Argentina. They are often engaged in bloody street battles in which they seek either to defend the honor of their club against other *barras bravas* or to control the economic resources of their own club.

The members of the *barras bravas* consider themselves to be the most fanatical and loyal members of the *hinchada*, that is, all the supporters that each sotccer team has. The passion and loyalty of the *barras bravas* is demonstrated on the venue of each soccer match, where they are found cheering for their team, even if attending the match means making long journeys across the country. Another duty that a *barra brava* must fulfill is defending the club's territory from any rivals: each club's stadium is located in the heart of these territories, also known as *barrios* (neighborhoods).

Some of the *barras bravas* hold long-standing and well-known rivalries, as is the case between Boca Junior's and River Plate's *barras bravas*, "La 12" and "Los Borrachos del Tablón". However, just as there are clashes between *barras*, there are also some strategic alliances that are forged with the sole objective of defeating and confronting common enemies. This is a phenomenon that occurs particularly among groups of supporters who are geographically distanced – they "become partners" in order to mutually support and/or defend themselves in their battles against a common enemy. This phenomenon can be studied by means of a multiple social network approach using the FANMOD software, available at http://theinf1.informatik.uni-jena.de/~wernicke/motifs/

The data was collected over a two-year period that combined field work and web research (from 2009 to 2011) at the University of Buenos Aires, and it consists of three kinds of network ties: rivalries among and alliances between *barras bravas*, as well as the number of games played by their football teams. There is also a set of attributes that includes the location (province and city) of the stadiums of each football club.

Our hypothesis is that the existing competitiveness in soccer and the geographic proximity of *barras bravas* are the elements which foster rivalries. The idea that sports competitiveness actively promotes social conflict has been widely documented in anthropologist literature, and such hypothesis can be tested following an SNA approach. As a matter of fact, it has been observed that there is an existing correlation between the number of games played by the soccer teams and the rivalry ties against their *barras bravas*, as well as an elevated homophily between those barras bravas that belong to the same province or city.

Given the fact that the ties of the Rivalry Network depend on the existing ties of the Game Network, and considering that the Game Network is a small world, it follows that the Rivalry Network is also small and clustered. The Rivalry Network is the first known empirical negative network that exhibits small world properties.

2. Data Collection

The dataset consists of 247 nodes that represent the most widely known *barras bravas* in Argentina. In these networks, the nodes stand for social groups, not individuals, so all the network ties are relations between groups, just like the international relations established between countries in a global network.

The data concerning the rivalries, alliances and attributes of each of the *barras bravas* were collected by means of field work and from the website www.barras-bravas.com.ar (no longer available). The rivalries data and the alliances data were symmetrized through the minimum method so all ties are reciprocal. In most of the cases, only one individual identified an existing rivalry or alliance between his own group and others, so there are limitations to the validity of the data. However, the symmetrization of the data is expected to partially solve this problem.

The number of games played was collected from the Rec.Sport.Soccer Statistic Foundation (available at http://www.rsssf.com). This data represents the number of games played between the *barras bravas*’s teams over the past 10 years. The soccer tournaments in Argentina are divided into five different levels or categories (First A, National B, First B/Argentine Tournament A, First C/ Argentine Tournament B, First D/Argentine Tournament C), and two geographic zones (the Metropolitan area of Buenos Aires, and the rest of the Country). The number of movements from one division to another is rather limited – as a result, a high regional clusterization is to be expected.

3. Data Files and Formats

All the data are provided in one Excel Workbook, Barras_ Dataset.xlsx, which consists of 4 worksheets (tabs). Three worksheets (Rivalries, Alliances, Games) contain the relational data, which are provided in a full matrix format. The attributive data (Barras_attr) is provided in a nodelist matrix format.

- Rivalries. Binary and symmetric data. The value of a cell represents the existence of a reciprocal rivalry.
- Alliances. Binary and symmetric data. The value of a cell represents the existence of a reciprocal alliance.
- Games. Valued and symmetric data. The value of a cell represents the number of games played among teams.
- Barras_attr. The following four attributes are provided:
    - o Name. The *barra brava*’s name.
    - o Club. The *barra brava*’s club. Useful as node label.
    - o State. The state where the stadium is located.
    - o City. The city where the stadium is located.

4. Data Details	
Response Rate	100%
Non-Respondent Bias	N/A
Theoretical Grouping	Negative Networks, Multiple Networks.
Publications Using These Data	• Bundio, JS. 2013. Redes Negativas: el Pequeño Mundo de las Hinchadas de Fútbol. REDES – Revista Hispana para el Análisis de Redes Sociales, • Bundio, JS. 2012a. El enemigo de mi enemigo es mi amigo. Explorando los conflictos y las alianzas entre hinchadas de fútbol. Lecturas, Educación Física y Deportes, 17(167), pp. 1-1. • Bundio, JS. 2012b. El Pequeño Mundo de las Hinchadas de Fútbol. In N. Kuperszmit, L. Mucciolo, T. Lagos Mármol & M. Sacchi (Ed.), Entre Pasados y Presentes III: Estudios Contemporáneos en Ciencias Antropológicas (pp. 236-246). Buenos Aires: INAPL. • Bundio, JS. 2011. Conflictos y alianzas entre hinchadas argentinas: apuntes metodológicos para el testeo de hipótesis mediante Análisis de Redes Sociales. Lecturas, Educación Física y Deportes, 16(155), pp. 1-1.
Data Context	Long-term anthropological investigation at University of Buenos Aires.
Respondents	Members of the barras bravas and soccer club’s followers.
Nodes	The soccer supporter's groups known as barras bravas.
Edges	Rivalries and alliances between barras bravas and number of games played among their teams
Longitudinal	None
Temporality	Medium. The barras bravas network are continually evolve, so the validity of the data will attenuate over time, especially in the medium-long term.
Analytical or Pedagogical Utility	• Illustrating homophily with categorical attributes. • Illustrating small world properties in both, positive and negative networks. • Illustrating the QAP Correlation.
Known Issues	None

Portland West Time Dollar Exchange Dataset

Ed Collom
University of Southern Maine
Portland, Maine

Author
Ed Collom, is a Professor of Sociology at the University of Southern Maine in Portland, Maine.

Please send all correspondence to collom@usm.maine.edu.

1. Overview

The Portland West Time Dollar Exchange (PWTDE) dataset contains the recorded transactions from a local currency group that existed in Portland, Maine for over four years. Such voluntary organizations allow participants to exchange services and goods without the use of federal money (see Collom, Lasker, and Kyriacou 2012). Unlike bartering (a direct swap between two parties), local currencies create a network of people and organizations in which transactions are tracked with an alternative currency. Time banks use time as their currency. The amount of time that a member spends helping another is entered in a database so that the provider is credited with "time dollars" (or "hours") and the recipient's account is debited. The other major form of local currency in the United States, the Ithaca Hours model, employs printed bills that members exchange for services or goods (see Collom 2005).

The PWTDE began in February, 2002 and was embedded in Portland West, a community-based social service agency. The organization ran out of grant funds to support its community outreach programs and was forced to close the time bank in June 2006 (Doherty 2006). At that time, all PWTDE members were invited to join Portland's larger and better-known time bank, the Hour Exchange Portland (see Collom et al. 2012).

The data consist of the 2,316 recorded transactions involving 6,712 hours of services exchanged among the 319 members at PWTDE over the course of its history. A multitude of social network analyses are possible with this dataset. It is longitudinal, directed, and valued. The date of each transaction is included, making it possible to investigate the evolution of the network across time (see analyses by quarter in Collom 2012). The ties are directed; one member has provided a service to another. The amount of time that the exchange took (the number of time dollars earned) is the value of the tie. Moreover, investigations of qualitative aspects of the ties are also possible as the services exchanged in the transactions have been categorized into 13 broad types (see Collom 2012; Collom et al. 2012):

> 1) Health and Wellness (e.g., yoga, acupuncture, meditation), 2) Beauty and Spa (haircut, massage, facial), 3) Office and Administrative Support (clerical help, bulk mailing), 4) Computers and Technology (computer repair, website design, audio/video production), 5) Tutoring, Consultation and Personal Services (lessons, tutoring, basic computer assistance, childcare), 6) Construction, Installation, Maintenance and Repair (carpentry, painting, yard/garden maintenance), 7) Cleaning, Light Tasks and Errands (cleaning, mending and alterations, errands), 8) Food Preparation and Service (cooking, catering), 9) Transportation and Moving (transportation, moving assistance, hauling), 10) Entertainment and Social Contact (companionship, performances, telephone assurance), 11) Events and Program Support (assistance with project/event, committee meetings), 12) Sales and Rentals of Items (purchase of used goods, space rental), and 13) Arts and Crafts Production (arts and crafts, artwork).

The dataset also includes three attributes of members. The first identifies whether the member is an individual or an organization. Most time banks have organizational members (usually nonprofits, community agencies, or small businesses). The gender and age of individual members are the other attributes.

2. Data Collection

This is a secondary dataset of information originally collected by PWTDE staff in a Microsoft Access-based software program called TimeKeeper (Gordon 1995). This software was designed for use by time bank coordinators to record member's hours of exchanges. In the early era of time banking, providers of services were instructed to contact the office (via phone, mail, or email) whenever they provided a service to someone. A staff member would then enter the transaction (including the provider, recipient, date of exchange, number of hours, and type of service) into the TimeKeeper database. Today, most time banks have web-based software in which members enter their own transactions.

The author was provided with an exported spreadsheet of relevant fields from the transaction table as well as the member table from the TimeKeeper database. With the exception of two organizational accounts identified below, the former members of this time bank are anonymous in these data.

3. Data Files and Formats

This dataset is provided in one Excel spreadsheet ("PWTDE.xls") which contains two worksheets (tabs). The first worksheet ("Transactions") contains the relational data. Each row represents a single transaction and lists the provider's ID, the receiver's ID, the length of time of the transaction (number of time dollars earned), the service category (described above), and the date of the transaction. The first transaction in PWTDE occurred on February 7, 2002 and the last on June 1, 2006. This relational data is complete with the exception that 95 of the 2,316 transactions (4.1%) are missing service categories and have been coded "99" (these are transactions that had been entered into TimeKeeper as "miscellaneous").

If users are employing UCINET 6 software (Borgatti, Everett, and Freeman 2002), the "edgelist1" or "edgearray1" DL file formats are most conducive for importing these transaction data. The latter allows one to import all of the attributes of the ties at once (note that the date field should be reformatted into a numeric value instead of the MM/DD/YYYY format). However, it will be easiest to just focus on the transactions (ignoring the service categories and dates) to begin. The edgelist1 DL file format can be employed to import the provider, receiver, and TDs (time dollars) fields. The values in the resulting matrix will be the sum of the total hours provided by each member to each other member. In other words, if member X provided a one hour service to member Y on ten separate occasions, the value of the X:Y tie in the matrix will be 10. If one imports only the provider and receiver fields, the resulting matrix will identify the total number of transactions X provided to Y. It is important to understand and recognize the distinction between the number of hours of services provided versus the number of transactions provided. The former is likely the more

4. Data Details	
Response Rate	100%
Non-Respondent Bias	N/A
Theoretical Grounding	N/A with the exception of the coding of the service types described above.
Publications Using These Data	Collom (2012); See Collom et al. (2012) for examples of analyses employing similar data.
Data Context	Recorded transactions from the database of a time bank that existed in Portland, Maine from 2002-2006.
Respondents	N/A
Longitudinal	Yes, the date of each exchange is included.
Temporality	Time banks vary dramatically. This dataset captures the complete history of one of these voluntary organizations.
Analytic or Pedagogical Utility	This dataset is all about who exchanges what with whom within a local currency group. See Collom (2012) for a list of key indicators of time bank participation that can be derived from this dataset. Additionally, a wide variety of social network concepts can be investigated with this longitudinal, directed, and valued transaction data and the accompanying member attributes.
Known Issues	Minor levels of missing cases and some time bank members do not report all of their transactions (see above).
Analytical or Pedagogical Utility	• Illustrating homophily with categorical attributes. • Illustrating small world properties in both, positive and negative networks. • Illustrating the QAP Correlation.
Known Issues	None

important since it captures the total time one member spent providing services to another. Of course, one could ignore both the number of hours and transactions and dichotomize the matrix to simply indicate whether X has ever provided a service to Y.

The second worksheet in the Excel file ("Attributes") contains the available attribute data on the members. The ORG variable distinguishes the 300 individual members (coded "0") from the 19 organizational members (coded "1"). Two of these organizational members are noteworthy. First, the time bank itself has an account since members often provide services for their time bank (such as clerical help in the office). In this dataset, PWTDE is member #2932. Portland West, the host agency of PWTDE, is member #2541 and the recipient of more services than any other participant (most of these hours were earned by people who volunteered in their learning center).

The SEX variable differentiates the 89 male members (coded "1") from the 211 female members (coded "2"). Such gender disparity is typical in time banks (see Collom et al. 2012). The AGE variable provides each individual member's age at the time of their initial participation (their first transaction). There were 22 members (7.3%) for which no birth date was available in the database (these missing values are coded "998" while the organizational members are coded "999" on this variable).

In addition to the missing cases involving the service categories and age attribute, it should be noted that transaction records are not perfect (see Seyfang 2001; Lasker et al. 2011; Collom et al. 2012). Some members do not report all of their transactions. One of the ironies is that unreported hours are sometimes the result of the success of time banking itself. As members get to know each other better and establish relationships with those with whom they are exchanging, recording transactions with friends may begin to seem unnecessary or even inappropriate. In other cases, members have high balances and simply do not bother. Some may also forget to report the services they have provided. Thus, while a time bank's transaction records reflect its "official" balances, they are an undercount of the exchanges that occur among members. It is not possible to know how such underreporting might bias this dataset.

References

Borgatti, Stephen P., Martin G. Everett, and Linton C. Freeman. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Collom, Ed. (2005). "Community Currency in the United States: The Social Environments in which it Emerges and Survives." *Environment and Planning* A 37 (9): 1565-1587.

. (2012). "Key Indicators of Time Bank Participation: Using Transaction Data for Evaluation," *International Journal of Community Currency Research*, 16: A18-29. Available online: http://ijccr.files.wordpress.com/2012/04/ijccr-2012-collom.pdf

Collom, Ed, Judith N. Lasker, and Corinne Kyriacou. (2012). *Equal Time, Equal Value: Community Currencies and Time Banking in the US*. Surrey, England: Ashgate Publishing.

Doherty, Nick. (2006). "Money Talks: Searching for the Almighty Time Dollar." *The Portland Phoenix* 6 July. Available online: http://portland.thephoenix.com/news/16762-money-talks/

Gordon, E. Kent. (1995). Timekeeper for Time Dollars Version 2.11.

Lasker, Judith, Ed Collom, Tara Bealer, Erin Niclaus, Jessica Young, Zane Kratzer, Lauren

Baldasari, Ethan Kramer, Rachel Mandeville, Julia Schulman, Danielle Suchow, Abby

Letcher, Anne Rogers, and Kathy Perlow. (2011). "Time Banking and Health: The Role of a Community Currency Organization in Enhancing Well-Being." *Health Promotion Practice* 12 (1): 102-115.

Seyfang, Gill. (2001). "Community Currencies: Small Change for a Green Economy." *Environment and Planning* A 33 (6): 975-96.

The Sunbelt 2013 Data: Mapping the Field of Social Network Analysis

Jürgen Pfeffer
Carnegie Mellon University,
Pittsburgh, Pennsylvania

Betina Hollstein
University of Bremen
Bremen, Germany

John Skvoretz
University of South Florida
Tampa, Florida

Please send all correspondence to jpfeffer@cs.cmu.edu.

1. Overview

The Sunbelt Conference is the annual meeting of the International Network for Social Network Analysis (INSNA). In May 2013 (5/22–5/26), the 33rd Sunbelt Conference was held in Hamburg, organized by Betina Hollstein, Sonja Drobnič, and Michael Schnegg. The keynote address (Simmel Award) was delivered by John F. Padgett ("Networks and History"). The Freeman Award lecture was presented by David Schaefer ("Distinguishing Pattern from Process: Equifinality and Network Selection"). 1,050 researchers from 50 countries participated at the XXXIII Sunbelt Conference to watch 656 paper presentations in 127 sessions (see network visualization in Figure 1) and 93 posters – overall 749 research products from 1,351 different authors.

Looking at the number of participants as well as the number of presentations, Sunbelt XXXIII was the largest conference in INSNA history so far. Participants represent a diverse array of scientific fields and cover a broad variety of topics. Consequently, analyzing the content of the conference can provide us with deep insights into the state of the art of research in the field of social network analysis.

This article describes the dataset that represents the printed conference program (including abstracts and keywords). We also describe the process of creating the program from the abstract submission system to the final placement of paper presentations to sessions.

2. Data Collection

For collecting the submissions of the paper and poster presentations, we turned to a market leader from Germany in the conference administration field, "pharma service" and used their online tool "Full Service Abstract management".

The submission system closed (after being extended for one week) on January 7, 2013. 848 abstracts have been submitted by researchers from around the world. This number includes a small number of abstracts that have been submitted slightly too late due to technical problems. During the submission process a set of information (authors, institutions, abstract, keywords, etc.) has been collected. These variables will be described in Section 3. We start out with the process of assigning paper presentations to sessions (time slots), as this is the main challenge of creating a Sunbelt program.

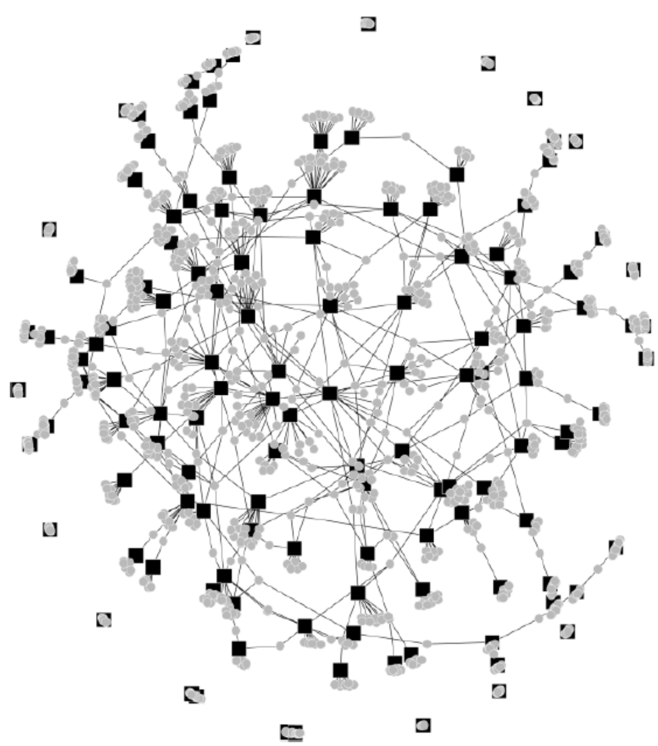


Figure 1: Sunbelt 2013 conference - sessions and people

2.1. Assigning Paper Presentations to Sessions

During the submission process, every person submitting an abstract was asked to select a preferred presentation form, namely either a 20 minute paper presentation or poster presentation; a third option was "no preferences". For the majority of submissions, the paper or poster selection is identical with the final presentation form; some who indicated the poster option in the special note section of the submission form have been altered accordingly by the local organizers in order to get hold of the large number of submissions. Most submissions indicating "no preference" were assigned to the poster session for the same reason. In this article, we primarily discuss paper presentations. Although, poster presentations are also included in the data (see next section).

For a paper presentation, submitters were asked to suggest a possible session topic[1] for the talk. A pre-defined selection of 72 session topics was compiled for this purpose based on sessions held at the last three Sunbelt conferences. However, it was also possible to create a new suggested session topic.

32 of the pre-defined session topics were organized, i.e., researchers had agreed on assisting in the session assignment process as well as on chairing the respective sessions at the conference. Most of the organized session topics had separate call for papers via the SOCNET email list and the conference website. 412 (=48.6%) of the submissions were covered by organized session topics.

It is important to notice that these suggested session topics are not always identical with the final session assignment. During the program creation process, some paper presentations were moved from organized to other session topics. Finally, session topics with lots of submissions were split to multiple sessions, and session topics with few submissions were joined. Beside the goal of putting similar abstracts in one session, there is another major constraint for the session assignment procedure. In general, the number of talks per session is fixed to five or six, except for a small number of sessions (normally the first session on the first day and the last session on the last day of the conference) less paper presentations are possible.

2.2. Rejects and Withdrawals

From 848 initial submissions 99 (11.7%) have been removed in the time period from January 7 to April 27, 2013. There were three reasons for a submission not getting into the final program:

- Rejected by the Sunbelt 2013 organizers if the submission did not comply with the minimum requirements[2] and the authors did not provide a re-casted version of the abstract after being asked to do so. (9)
- Rejected because of missing conference registration (28)
- Withdrawn by the authors (62)

This Sunbelt 2013 dataset reflects the state of the printed program (deadline April 27, 2013) including 656 paper presentations in 127 sessions and 93 posters. Later withdrawals (that have been removed from the interactive online program3) remain in this dataset.

3. Data Files and Formats

The Sunbelt 2013 data consists of several tables. The tables are stored in separate sheets of a single Excel file ("SUNBELT 2013 Data.xlsx"). These tables were used to create the printed and the interactive online program. There are two key variables that connect the tables and both are essential to the program:

ID: A unique identifier for every submission. No overlap between paper and poster presentations. IDs connect authors, titles, affiliations, keywords, and abstracts in different data tables.

Session Code: A code that connects a presentation with a specific session. These codes are constructed from four parts, the weekday, morning or afternoon, first or second session (see sheet "TimesSlots"), and the room ID.

4. Data Details

In the following, we describe the data tables of the Sunbelt 2013 data. In general, data columns are in original form as provided by the researchers when submitting their abstracts. Post-processed data are reported explicitly. Please notice that for both paper and poster presentations

[1]We use the term "session topic" to take into account that some of which resulted in multiple sessions.
[2]Sunbelt XXXIII Conference Guidelines & FAQs, Abstract Submission FAQ, page 9: "The abstract must describe some work that is about NETWORKS, most likely, social networks..."
[3]www.insna.org/program2013

a one-presentation-per-person rule were carried out and that the first author is not necessarily the presenter.

4.1 Paper Presentations

This table includes all 656 paper presentations. Every data line represents one submission. Here are the columns of the table:

ID	Submission identifier
Abstract Title	Title of the paper presentation
Author(s)	Author(s) of the submission. The presenter is underlined; superscript numbers connect people to institutions in case there is more than one institution involved.
Institution(s)	Institution(s) of the author(s); superscript numbers connect to author(s).
Country	Country of the person that submitted the abstract. The person doing the submission is not necessarily the presenter or the first author.
Session Title	Title of the session in which the paper was presented. This is the assigned session (see section 2.2.) not the session topic suggested by the author(s).
Session Code	Day/Time slot and room ID describing when and where the paper was presented.
Talk Nr	One session consists of multiple talks (normally five or six). This number indicates the position of the paper presentation within the session.

4.2. Poster Presentations

This table includes all 93 poster presentations. Every data line represents one submission. The columns in this table are identical to those described in the previous section (paper presentations) except that there are no sessions assigned to posters as all posters have been presented in a single poster session.

4.3. Keywords

This table shows the selection of keywords by the people submitting the abstracts. Every keyword/submission link is a single line to make it easier to import this table into a network tool. The selection of keywords was restricted in the submission system by two factors. First, the keywords were pre-defined and no new keywords could be entered. A list of 100 keywords was compiled from keywords used for the last three Sunbelt conferences. Note that there is also a keyword “others”. Second, the number of possible keywords to select ranges from 1 to 5. 2,813 keywords have been selected (avg. 3.76) for all 749 paper and poster presentations. Every single keyword was used at least two times. The top used keywords are Social Capital, Egocentric Networks, and Inter-organiza-tional Networks. The columns of the keywords table are defined as follows.

ID	Submission identifier
Type	Paper or poster presentation
Keyword	Keyword selected from a pre-defined list of keywords

4.4. Institutions

For further post-processing of the authors and institutions outside of Excel, the superscript affiliation numbers can turn to an intricate problem. This table should ease this data manipulation pain. Still, in order to get your institutional collaboration network, some cleaning will be necessary as these are text data that were typed separately for every submission. The columns of this table are copies of previously described tables.

4.5. Index

This table was created for the index of the printed program. This table (excluding the posters) was also the data source for the Sunbelt 2013 conference poster of sessions and people (Figure 1).

Every line connects one person with a specific session represented by the Session Code. The names in this table have been cleaned so that different writings of one person result in a single name, e.g. “Pattison, Philippa”, “Pattison, Philippa E.”, and “Pattison, Pip” were all converted to “Pattison, Philippa E. (Pip)”. This cleaning process was just done for this table and did not change the writing of the authors in other tables. Note that all posters have been coded with “Poster”.

Name	Author of a submission (cleaned).
Session Code	Day, time slot, and room ID of the session in which the author gives a paper presentation

4.6. Abstracts

749 abstracts are part of the printed program. Every abstract can be found in a single line in this table. 157,000 words of these abstracts added up to the 300 pages “Abstract Program” that can be found online at the conference website. Be aware that some researchers like to prepare their abstracts in MS Word or other text processing tool and then copy/paste the abstract—including formatting

and special symbols—to the abstract submission system. If you are planning to analyze the abstracts, be aware that most of the original formatting (including paragraphs) has been deleted in the data handling process; sometimes by leaving enumeration symbols in the middle of the text, e.g., see submission 457. Also pay attention to the fact that older Excel versions crop the abstracts due to a limitation to the number of characters by cell.

ID	Submission identifier
Type	Paper or poster presentation
Abstract	Abstract of the submission

4.7. Sessions

The session table describes when and where a session was held and if that session had organizers (see section 2.2). In case of an organized session, the organizers served as session chairs, otherwise the last presenter of the session was assigned as session chair. The maximum number of talks per session is six. Multiple sessions that are the result of a single session topic (see section 2.2) can have enumerations (e.g., Social Capital 1, Social Capital 2) or title extensions (e.g., Words and Networks: Health and Culture, Words and Networks: Politics and Crises, …).

Session Code	Day, time slot, and room ID of the session
Session Title	Title of the session
Day	Day of the conference (Wednesday=1, Sunday=5)
Weekday	Wednesday, Thursday,...
Slot	Time slot of the session (see table *TimeSlot*)
Room	The ID of the room where the session took place.
Organizer(s)	Name(s) of the organizer(s) in case of organized sessions

4.8. Times Slots & Paper Presentation Times

In order to create the program, two more tables were necessary. The table TimeSlots defines the time Slot entries of the tables described above and the table PresentationTime maps a given time Slot and a Talk Nr combination to the exact starting time of the paper presentation.

5. Additional Materials

A high-quality printable PDF version of the Sunbelt 2013 conference poster of sessions and people as well as the related network file can be found here: www.pfeffer.at/sunbelt2013

Acknowledgements

Several people have been engaged in creating this dataset. Dagmar Zanker compiled the keyword list as well as the list of suggested sessions for the registration system. Both lists are based on the work done by Laura Koehly, Rebecca Davis, and Tom Valente, the organizer of Sunbelt XXXII (2012) in Redondo Beach, California. Julie Hewett (INSNA) managed the registration system of the conference. Sonja Drobnič and Michael Schnegg reviewed the abstracts together with Betina Hollstein. 61 network scientists organized sessions and helped with assigning talks to sessions. Tom Töpfer helped clean the names for the index. Finally, thanks to the many people that actually wrote and submitted the 749 abstracts and presented their work at Sunbelt 2013 in Hamburg.

Brief Report

Knowing Your Social Network Data and Measures and Big Data: A Summary of Jeffrey C. Johnson's Keynote XXXIV Sunbelt Social Network Conference

Tracy Van Holt
East Carolina University
Greenville, North Carolina

Author
Tracy Van Holt, is an assistant professor in the Department of Geography at East Carolina University in Greenville, North Carolina.

Please send all correspondence to vanholtt@ecu.edu.

Electronic data from the internet and cellphones have made an unprecedented amount of information about human behavior and social structure available. The analysis of Big Data (BD) is increasingly playing a role in social network research. But while BD is appealing, there are trade-offs. With what would now be considered small data (SD), it is easier to test for reliability and validity because there is more control in the research design and data collection stage. With BD, which are often mined, it is harder to know if these data are meaningful because there is little control over the questions asked of respondents and there is no on-the-ground understanding of the kind that comes from direct interviewing or from participant observation. Jeff's keynote summarized how ethnographic knowing, aids the development of theory, of measures, and of data collection in the study of social networks and how ethnographic knowing can be applied to BD.

While writing a blog post about Jordan White (a singer-songwriter who was a "rising star" with a big following on Twitter), Emily Guendelsberger, found that the BD she was analyzing was filled with false leads. Based on a hunch, Guendelsberger spent a day tracking Jordan's Twitter followers and found that many of the 200,000 followers that Jordan had amassed were an artifact of mass behavior—an illusion. BD can be filled with false leads and requires ethnographic knowing.

For his doctoral research, Jeff immersed himself in an Alaskan Salmon Fishery as a fishing camp boat carpenter and did SD ethnographic research on scarcity and resource management. As a carpenter, Jeff saw how informal social roles affected group structure, dynamics, and outcomes. One person was willing to be the butt of jokes and pranks—a role labelled "court jester" (Johnson and Miller 1983). Academics at that time mostly had a negative view of social deviants like the court jester. Indeed, the court jester was the worst fisher in the camp and he had no kinship relations to the group. However, Jeff observed that the court jester fostered group cohesion when the fishers in the camp were stressed over a strike. He was rewarded for his role: this ostensibly lower status actor hung out with and interacted with higher status actors in the camp and he was rewarded for his informal role with over-the-limit fish transfers. Had Johnson not known his informants well, the court jester role would have appeared less influential in group dynamics.

Building on this work, at the Antarctic Research Station Jeff and colleagues James Boster and Lawrence Palinkas tested the hypothesis that networks that evolve into core-periphery structures—those with higher global coherence--will function better than groups that evolve into clique structures (Johnson et al. 2003a). Each year, for three years, they did ethnographic research on small groups—~30 people each—that lived in winter-over isolation at the station for eight months. The groups contained the researchers (the beakers) and the people who ran the station (the trades) from different genders and social classes. Over the course of the Austral winter, crew were given a series of ratings and pile sorting tasks on social interactions to map the networks. A globally coherent group structure, that is, a core periphery structure with no subgroups were identified in year 1. In year 2, a core group of people initially emerged and some cliques developed over the year, though the structure still maintained a high global coherence. Distinct clique structures evolved in year 3 and this network had low global coherence (Johnson et al. 2003a). The research team created an index of core-periphery structure [the coefficient of relative variation (CRV)] by incorporating the first factors from a factor analysis of the correlation matrices that represented the interaction among station members (Johnson et al. 2003a). Low CRV scores reflected lower variation in the first factor and a higher core-periphery structure, or higher global coherence (Johnson et al. 2003a). When he included multi-year data from research groups from stations in Poland, Russia, and China where each station-year represented a unit of analysis, higher morale was highly correlated with higher global coherence (Johnson et al. 2003b).

The team also tested whether formal and informal leadership roles (instrumental and expressive leadership) and the role of social deviants explained the evolution of the social network structure (Johnson et al. 2003a). Positive deviants (like the court jester) foster cohesion in times of stress, while negative deviants (like people who challenge the leadership) can be disruptive. Roles were measured by a sentence completion task, with questions like: "__________ is a natural leader in getting things done around the station" and " __________ is one of the entertainers or comedians." They examined the consensus on perceptions of the formal and informal leadership roles and deviants at the end of winter and compared that to global coherence. The networks in years 1 and 2, with more globally coherent structure showed a trend toward high agreement on and overlap of both formal and informal leadership roles compared with the network in year 3, with low coherence. Globally coherent structures (years 1 and 2) were also associated with positive deviants and lower coherence networks had fewer positive deviants and more negative deviants (year 3) (Johnson et al. 2003a).

Jeff's keynote address raised the question: Can we gain similar understanding of human behavior in BD analysis, where it is difficult to know the informants or to get ethnographically informed social measures of interest (e.g. trust)? Much of BD work is being driven by data and are exploratory in nature rather than being driven by theory. BD analyses search for patterns or key structures or nodes. But is there theoretical meaning to the metrics used? Are there theoretical and empirical meaningfulness of these patterns? How are these patterns associated with outcomes? Knowing in BD is the big challenge. If Jeff just had the networks from the South Pole station without the ethnography he could not have understood the social structure or developed a theory to account for the structure. The equivalent in BD is to find original sources and check for validity and reliability. Johnson, Van Holt, and their collaborators have been applying ethnographic knowing to BD. They compared the trade-offs in recall and accuracy between human-coded and machine-coded datasets (Van Holt et al. 2013) and they tested whether environmental factors were linked to conflict as theory predicts by analyzing eight years of articles from the Sudan Tribune; they show that ethnic groups with more news reports of severe conflict had more news links to livestock (Van Holt et al. 2012).

Johnson will be working on the sequel to this talk and the search for meaning in mined data in his new position as a Preeminent Professor of Anthropology at the University of Florida's initiative on informatics and big data.

References

Johnson J.C. and M.L. Miller. 1983. Deviant Social Positions in Small Groups: the Relation Between Role and Individual- *Social Networks* 5:51-69, 1983.

Johnson, J.C., Boster, J.S. and L. A. Palinkas. 2003a. Social roles and the evolution of networks in extreme and isolate environments. *Journal of Mathematical Sociology* 27 (89-121).

Johnson, J.C., Palinkas, L.A, and J.S. Boster. 2003b. Informal social roles and the evolution and stability of social networks. In "Dynamic Social Network Modeling and Analysis". The National Academies Press: Washington, D.C.

Van Holt, T., Johnson, J.C., Brinkley, J., Carley, K., and J. Diesner. 2013. Rapid Ethnographic Assessment for Cultural Mapping. *Poetics* 4(4): 366–383

Van Holt, T., Johnson, J.C., Brinkley, J., Carley, K.M., and J. Caspersen. 2012. Structure of ethnic violence in Sudan: an automated content, meta-network, and geospatial analytical approach. Computational and Mathematical Organization Theory 18:340-355.

Book Review

John Scott, **Social Network Analysis**, Third Edition. Los Angeles, London and Thousand Oaks, CA: SAGE Publications, 2013, pp. x+201.

John Scott, **What Is Social Network Analysis?** London and New York: Bloomsbury Publishing, 2012, pp. x+126.

Ronald Breiger
University of Arizona
Tucson, Arizona

Author
Ronald Breiger, is a professor in the School of Sociology at the University of Arizona in Tucson, Arizona.

Please send all correspondence to Breiger@email.Arizona.edu

John Scott has contributed two recent sole-authored volumes introducing social network analysis. One is a welcomed third edition of his bestselling text, and the other is a new and highly compelling introductory guide within Bloomsbury Academic's "What Is" research methods series (with other volumes already or soon to be published on topics including online research, qualitative research, community studies, and narrative research). I will abbreviate these new volumes by John Scott as SNA3 and as WhatIs, respectively.

Researchers and students, including college undergraduates, from diverse fields who have very little knowledge of social network analysis and who do not have or seek a strong background in technical-mathematical ideas will find in these volumes a spirited motivation for undertaking social network research, in the form of clear presentations of key concepts and, of special importance, well-crafted reviews of research studies in which the author shows how network thinking and methods have made a difference to the understanding of social and economic life. These are characteristics of both volumes under review. These volumes differ primarily in the following ways. SNA3 is largely focused on illustrations of key concepts and metrics, featuring a total of some 55 Figures (for example, a graph on 16 nodes accompanied by a table showing each node's degree, relative degree centrality, and sum of geodesics; a table of illustrative data for nonmetric multidimensional scaling showing an original matrix of dissimilarities, an ordinal transformation, and loadings on a two-dimensional solution; a series of blockmodel images illustrating a variety of hypothesized power relations). By way of contrast, the WhatIs volume largely avoids this degree of elaboration (this latter volume has only eight Figures). However, this reviewer was highly impressed by the clear and even cozy style of a volume that, more than a textbook or a handbook could do, draws the novice reader in to a lively introduction to the history, key concepts and measures, applications, and even an excellent section on criticisms of social network analysis.

The first edition of SNA3 was published in 1991 (as Social Network Analysis: A Handbook), and a second edition (with the same subtitle) in 2000. SNA3 adds a new chapter on network dynamics, and the author states (p. x) that he has completely updated the references and discussions. The author's aim has been "to simplify the techniques" of social network analysis so as to make this family of activities "accessible to those with a limited mathematical background," and he hopes that there is enough "to satisfy both the newcomer and the more advanced researcher in need of information on current techniques" (p. x). As a noted authority and contributor to social network analysis, the author serves both newcomer and seasoned researcher with an admirably broad range of literature citations to encourage more learning in depth. An introductory chapter ("This book is a guide or handbook…, and not a text to be read through at one sitting," p. 9) is followed by a core of two chapters that the newcomer is advised to read first (chapters on history and on research design), followed by six chapters on the specifics of network analysis, ranging from density (illustrated by a review of Wellman's studies of communities) to centrality (made vivid by a review of work by Schwartz, Mariolis and others on bank centrality in corporate networks) to positions, sets, clusters, and blockmodels (exemplied by studies by Burt and by John Scott on corporate interlocks and participations in the U.S., Britain, and Japan). Other topics treated at length include cliques, cores, and components; positions, sets, clusters, and blockmodels; dimensions and visual displays of network data; and the newly written chapter on network dynamics, the modeling of change, and explanation testing.

Chapter 6, on components, cores, and cliques, is illustrative of the organization of each of the central chapters. The chapter opens with a reference to the role of the clique in early community and workplace studies (by W. Lloyd Warner and his associates and by Roethlisberger and Dickson, respectively), which had already been motivated in the earlier discussions of the history of this interdisciplinary field. Following discussion of the idea of a fully connected subgraph, connected components are introduced, illustrated in a diagram, and the point-and-click sequence is given for computing components using the Pajek program. Cut-points and Everett's EBLOC procedure are introduced, all without invoking any formulas or equations. Directed graphs (and semi-cycles) are similarly introduced. Nesting of components (k-cores, based on nodal degree, and m-cores, based on the multiplicities of arcs) are then discussed, based on the author's generalization of Seidman's work. Two successive figures illustrate (respectively) a 3-core and the collapse of a 3-core, based on application of Seidman's idea of the k-remainder; collapse of an m-core is illustrated in the following figure. Doreian's specification of the formal properties of a clique are introduced, as is Seidman and Foster's concept of the k-plex. Work by Kadushin, Alba, and Moore on the identification of social cirlces, shown as indebted to earlier theorization by Simmel, is reviewed. While the discussion throughout is non-mathematical, it is clear that, just as the novice benefits from an introduction to the concepts, ideas, and motivations, at the same time the advanced researcher is likely to be reminded of relevant previous work that deserves to be thought about anew. As do all the main chapters, this one ends with a review of research studies exemplifying the ideas (in this case, research of D. Crane, N.C. Mullins, A.C. Gattrell, M.E.J. Newman, and others on citation circles).

As with all worthy guidebooks, routes not taken might stimulate further discussion. Thus, while "The Harvard breakthrough" (a section of the history chapter) well satisfied this Harvard-trained reviewer, it is nonetheless a striking omission that research of other centers (notably that of the University of California at Irvine, which of course produced work that suffuses this or any other review of social network analysis; the University of Chicago [though Laumann's PhD is indeed from Harvard, and his work is discussed in the section on community elites]; and earlier the Columbia University of political influence, mass media, and national elite studies) is essentially ignored in the chapter on history and breakthrough. Work of Doreian, Batagelj, Ferligoj, Mrvar, and their colleagues that has led to breakthroughs in generalized blockmodeling is barely mentioned (pp. 135-6) in the blockmodeling chapter. Likewise, work on ERGM models is given just three paragraphs (pp. 144-45) in the new chapter on network dynamics, where the author's decision to avoid math seems especially constraining, even though he (I think correctly) characterizes "the move to dynamic models, longitudinal analysis, and significance testing" as "perhaps, the most important advance … since the Harvard innovations of the 1960s and 1970s" (p. 145). This volume is also essentially silent on the relatively new work on social networks and culture, and on geospatial analyses. The chapter on visualization could have benefitted by incorporation of insights from Linton Freeman's chapter on that topic in a 2005 volume co-edited by the author (P.J. Carrington, J. Scott, S. Wasserman, eds. *Models and Methods in Social Network Analysis,* Cambridge).

As mentioned, the other volume reviewed here, WhatIs, lacks the degreee of elaboration that I have discussed and illustrated above. It nonetheless has important merits. It is written in a highly compelling, conversational style and maintains the voice of an author and engaged researcher rather than the more neutral tones of the guide or handbook writer. (In the old days, as recounted on p. 5, the computer demands for running Clyde Mitchell's version of the CONCOR program "were so great that it could run only because my university rented time on the Manchester super-computer: which was almost as powerful as one of today's mobile phones!") A brief introductory chapter is followed by chapters on history (organized now according to the broad mathematical approaches that are said to have dominated the field: graph theory, algebraic, and spatial), key concepts and measures, applications of network analysis (including Moody and Light on citation analysis of the overall structure of sociology), an especially welcomed chapter on criticisms and frequently asked questions about social network analysis, and a brief chapter on software that mentions UCINET, Pajek, and some of the packages written in R. The criticisms of social network analysis that are discussed (along with responses) include: representations of work as innovative new even though a reinvention of already-existing procedures due to lack of historical knowledge; the charge of triviality; the charge of doing work that is unnecessary; the claim that it's just pretty pictures; the claim that it's simply too formal; and the claim that it's static. The frequently asked questions (to which responses are provided) include, among others: How can I decide who to include as members of my network? Networks include positive as well as negative relations; does this pose any problems for analysis? What are the ethics of social network analysis; isn't it just a form of snooping and spying?

Both volumes reviewed here provide strong introductions to social network analysis that will be of great interest to novices and that also provide benefits to instructors and advanced researchers. Despite some unavoidable overlap, the volumes complement each other well, and both should find a place on the shelf of many students, teachers, and researchers.

Book Review

McCulloh, I., Armstrong, H., & Johnson, A. **Social Network Analysis with Applications**, John Wiley & Sons, 2013.

Jürgen Pfeffer
School of Computer Science,
Carnegie Mellon University,
Pittsburgh, Pennsylvania

Author
Jürgen Pfeffer, is a Assistant Research Professor at the Institute for Software Research in the School of Computer Science at Carnegie Mellon University in Pittsburgh, Pennsylvania.

Please send all correspondence to jpfeffer@cs.cmu.edu.

The textbook "Social Network Analysis with Applications" was written by Ian McCulloh (Research Fellow at Curtin University in Perth), Helen Armstrong (Associate Professor in the School of Information Systems at Curtin University in Perth), and Anthony Johnson (Associate Professor in the Department of Mathematical Science at the United States Military Academy at West Point) and was published by John Wiley & Sons in 2013. The book starts with a short but very nice and almost thrilling introduction to the history and the crucial personalities of social network analysis. The authors span their historical overview from more than hundred years ago with Comte and Simmel to the foundation of the International Network of Social Network Analysis (INSNA) and its driving actors.

The actual chapters are organized into four parts. The first part covers centrality metrics and network topologies. The main concept of this textbook is obvious from the first page. The authors show many examples, draw network visualizations, and show matrices and network calculations in great detail. The authors are convinced "that an understanding of the mathematics leads to deeper and more complete understanding of the social concepts behind social network analysis" (p. xx). Although, the first part of the book has many equations, the algorithms, the math, and the underlying social considerations are discussed in every possible detail.

Part two is called "social theory" covering social forces (e.g. homophily, reciprocity, transitivity), grouping, and diffusion. Here, the authors try to walk the tightrope of including many well-discussed SNA concepts as well as rather new approaches, e.g. link optimization. Part three "data" covers briefly data collection methods and some related issues (data quality, anonymity). The larger portion of this part gives an introduction to multi-modal analysis using the meta-network approach that has been developed by Kathleen M. Carley and David Krackhardt and that incorporates knowledge, resources, etc. into social networks. The authors also describe the matrix algebra that is necessary to manipulate these meta-networks.

The final part of the book is dedicated to organizational risk and targets practitioners. The authors give instructions on how to interpret the results of network metrics from the perspective of assessing organizational networks. They also discuss the impact of possible network interventions, e.g. adding or removing nodes or edges. Finally, the authors provide a guideline for the process of organizational risk analysis including possible questions that can be asked during a network assessment (pp. 228ff). This part of the book can be of special interest for consultants or for people doing network analysis within a company. The book closes with an introduction to matrix algebra and an enumeration of all network data used in the book's examples.

At the end of every chapter the authors provide lab exercises that show how to accomplish the calculations of the chapter. These exercises are step-by-step instructions (with screenshots) using the tool ORA (developed at Carnegie Mellon University) and can be used in class or for self-study. Checks on learning progress are sprinkled through the entire book; related answers are provided directly after the questions.

Compared to other recently published textbooks in the field of social network analysis the book by McCulloh at al. is easy to digest which makes “it especially appropriate for newcomers to the study of networks”, as Katharine Faust states in the foreword (p. xvii). The basic concepts of SNA (e.g. centrality metrics) are much more detailed described than in other SNA books. This makes it perfect for self-study or newcomers to SNA. The key characteristic of this book is its unpretentious approach to network analysis – the authors use SNA as a toolbox. For instance, one of the first lab exercises uses a homonym network of words that have letters in common without discussing issues related to networks from words. But this would not make any sense at this point anyway.

In contrast to applications and exercises, the authors do not spend much time on discussing theories or methods without directly discussing applications and use cases. A side effect of this approach is that some “big” SNA concepts like social capital and structural holes are mentioned in one paragraph each. However, the authors provide information about related literature for the interested reader. On the plus side, the very practical approach is a very efficient way to quickly get a sense of achievement, especially, if you are a newcomer to SNA and you need to accomplish a network analysis for your class work or for your company by the end of next month. In a nutshell, if you are a self-study beginner or an undergrad lecturer and find other recently published textbooks too hard for yourself or for your students to start with, then this is your book for the first part of the studies. For the advanced social network analysis researcher who is not so familiar with algorithmic details of commonly used metrics, this book might offer some additional insights.

INSNA Business Meeting Minutes

February 22, 2014 12:00pm - 1:40pm
St. Pete Beach, Florida, USA

Thanks to the St Pete Beach organizing team. The Wellman Award for Outstanding Tweet of the Meeting. Thanks to the editorial board and reviewers for Connections for all their hard work.

1. Some stats on Subelt XXXIV

Year	Number	Location	Registrations	Workshops	Enrollment
2014	XXXIV	St. Pete Beach, FL, USA	633	34	470

2. Organizer report on Sunbelt XXXV, Brighton UK, 23-28 June 2015

3. Report from Board Meeting

   a. Nominations for board members
   New or renewed for three year terms 2014-2017

   Ulrik Brandes, Ainhoa de Federico de la Rua, Garry Robins, Dimitris Christopoulos, David Lazer

   Continuing Board Members (terms ending in 2016)

   b. Approved the Sunbelt XXXV Organizers' selection of keynoter and Simmel Award recipient, Tom Valente.

   Yanji Bian, Katie Faust (Vice-President), Emmanuel Lazega, Carter Butts, Laura Koehly (Treasurer), Alessandro Lomi, Noshir Contractor, John Skvoretz (President)

   c. Board agenda items

   1. Clarified Sunbelt rotation agreed by board in 2012 on a 6 year cycle, three times in USA/North America, two times in Europe, once in nonEuropean/nonNorthAmerican location – Other, for short – if a suitable one can be found (otherwise in North America).
   2. Reviewed proposals for Sunbelts in Hawaii, Chicago, Beijing, and Utrecht. In light of the clarification of the above rotation and the board's desire for better understanding of financial responsibilities and commitments, it was decided to ask submitters to revise proposals to include detailed conference budget plans following a template to be developed by the Board's finance committee, Laura Koehly (Treasurer), Nosh Contractor, and Garry Robins, with cost fields (printing programs, purchasing and loading flash drives, etc.) prepopulated where possible from past experience.
   In this context, the board also discussed pros and cons of adjusting workshop fees and revenue sharing between presenters and INSNA, and the setting conference registration and membership fees. No final decisions were made.
   3. Asked the Publications Committee, chaired by Carter Butts, to assess the status and future direction of INSNA publications, JoSS and Connections, and the costs involved in alternatives and the cost impact on membership dues. Committee members include Ron Breiger, Dimitris Christopoulos, Jim Moody, Tom Valente.

   4. Discussed bylaw revisions particularly with respect to voting procedures, definition of quora and so on and the standards that apply given our incorporation as a 501c3 in Delaware. The president was charged researching the issues further.

International Network for Social Network Analysis

Connections is the official journal of the **International Network for Social Network Analysis** (INSNA). INSNA is a scientific organization made up of scholars across the world. Updated information about INSNA's annual conference (**Sunbelt Social Network Conferences**) can be found on the website at *www.insna.org*.

INSNA (International Network for Social Network Analysis)

Hardcopy circulation of *Connections* is sent to all members of INSNA, the International Network for Social Network Analysis, which has over 1300 members. Subscription to *Connections* can be obtained by registering for INSNA membership through the website: www.insna.org. Standard annual membership fee is US$120 ($80 for students). Wherever possible, items referenced in articles (such as data and software) are made available electronically through the INSNA website. In addition, the website provides access to a directory of members' email addresses, network datasets, software programs, and other electronically stored items.

Sunbelt Social Network Conferences

The Sunbelt Conferences bring researchers together from all over the world to share current theoretical, methodological, and empirical findings around social networks. While previous annual conferences for INSNA members have been held in North America and Europe, conference locations will now expand to include other countries such as Austalia and those in East Asia. Information on the annual Sunbelt Social Network Conferences can also be found on the INSNA website. Sunbelt XXXV will be held in Brighton, UK June 23-28, 2015.

Manuscript Submissions

Submit articles to editorconnections@gmail.com. Full manuscripts should be submitted as an MS Word document, and should not exceed 6000 words including all tables and figures Brief Reports submissions should be under 2500 words. All images, figures and tables should be sent as seperate files. Rastor (photographic) images and figures should be sent in a high resolution (300ppi min.) graphics format (EPS, TIFF, or JPEG), while line work images and figures should be sent as a vector-based format (EPS or SVG). Format and style of manuscript and references should conform to the conventions specified in the latest edition of the Publication Manual of the American Psychological Association. Further instruction on submission formatting can be found on the INSNA website, http://www.insna.org. Manuscripts that do not follow submissions criteria will be returned for revision. The journal follows a double-blind review process for research articles. (ISSN 0226-1776).

DEN (Data Exchange Network)

We recently introduced a new section in *Connections*, which functions as an outlet for publishing datasets. The Data Exchange Network (DEN) has been designed with two goals: first is to build a community resource for network datasets and the second is to provide a format for citable references of datasets and instruments. Submissions should include an electronic version of the network dataset and/or instrument and a short article (not to exceed 2500 words) describing the data being submitted. All materials submitted for the DEN will be peer-reviewed to ensure the utility and usability of the data/instrument. Accepted DEN contributions will appear in the hard copy of *Connections*, and the data sets will be available on the INSNA website through an indexed, searchable web interface.