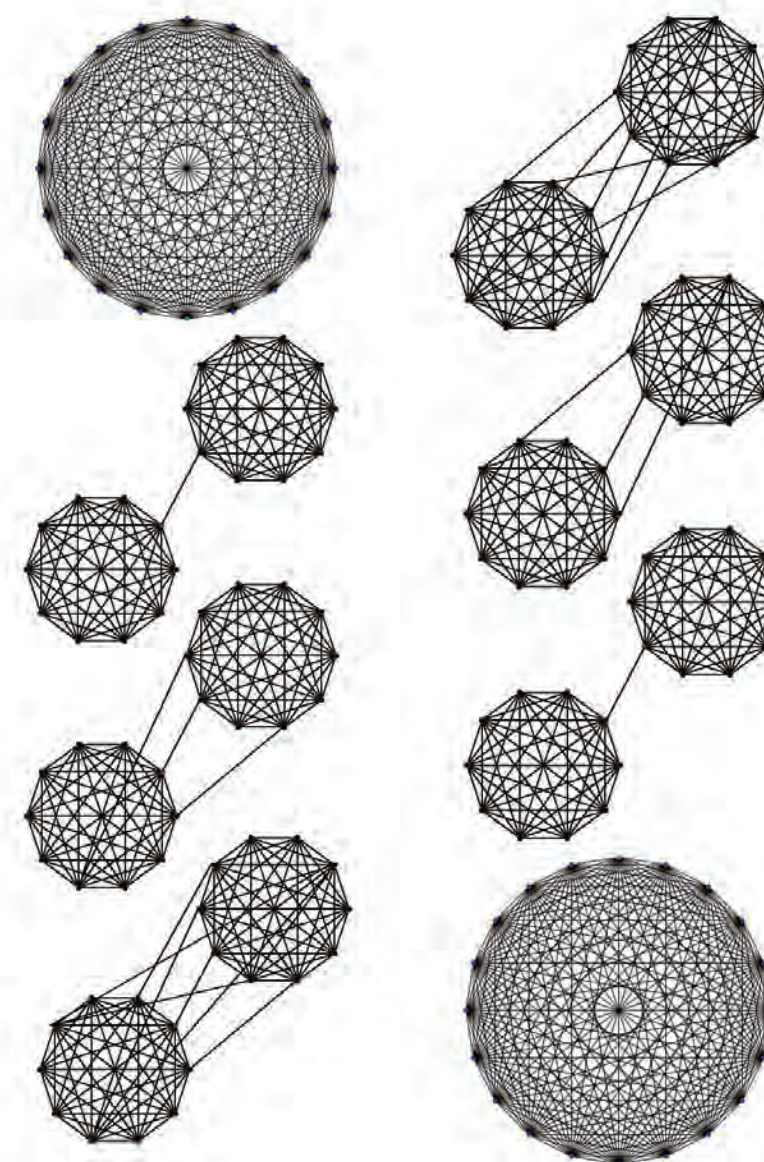


CONNECTIONS

December 2010

Volume 30 Issue 2

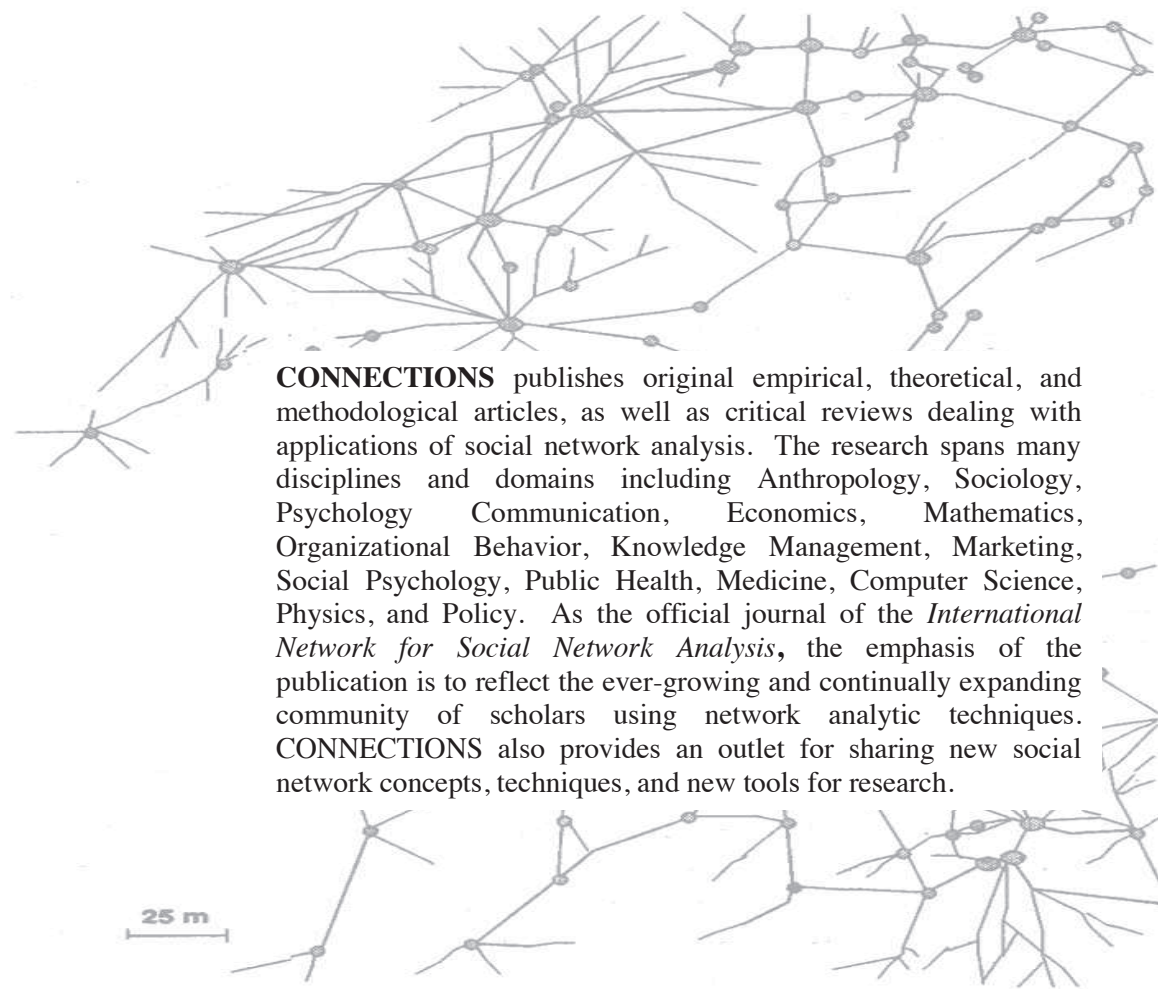


Official Journal of the International Network for Social Network Analysts

December, 2010

CONNECTIONS

Volume 30 • Issue 2



CONNECTIONS publishes original empirical, theoretical, and methodological articles, as well as critical reviews dealing with applications of social network analysis. The research spans many disciplines and domains including Anthropology, Sociology, Psychology, Communication, Economics, Mathematics, Organizational Behavior, Knowledge Management, Marketing, Social Psychology, Public Health, Medicine, Computer Science, Physics, and Policy. As the official journal of the *International Network for Social Network Analysis*, the emphasis of the publication is to reflect the ever-growing and continually expanding community of scholars using network analytic techniques. **CONNECTIONS** also provides an outlet for sharing new social network concepts, techniques, and new tools for research.

Front Cover: Image is from enclosed article titled "What You Believe Travels Differently: Information and Infection Dynamics Across Sub-Networks" by Grim, et al. The series of networks shown are clearly related in terms of structure. The larger network is a single total network, also known as a complete network or maximal graph. The three pairs of smaller networks form paired sub-networks with increasing numbers of connecting links. Whether dealing with a single network (large network) or sub-networks with increasing numbers of links (smaller networks), time to total infection across a network is most sensitive to network structure--ring, total, small world, random, or scale-free. The dynamics of information is importantly different. For information with reinforcement, it is not network type but degree of linkage across single or sub-networks that plays the major role.

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 the INSNA's annual conference (**Sunbelt Social Network Conferences**) can be found on the website at www.insna.org.

INSNA Board Members

President: George A. Barnett
 Vice President: Pip Pattison
 Treasurer: Thomas W Valente
 Founder: Barry Wellman
 Members: Phil Bonacich, Martin Everett, Katie Faust, Scott Feld, Anuska Ferligoj, Garry Robins, Ulrik Brandes, & David Lazer

INSNA Committees

Finance Committee - Chaired by Treasurer Tom Valente

Conference Committee - Chaired by Sunbelt hosts Russell Bernard and John Skvoretz

Web Committee - Chaired by webmaster and Chief Information Officer Benjamin Elbirt

Publications Committee - Chair - TBD

Composed of current and former editors of *Social Networks*, *CONNECTIONS* and *Journal of Social Structure (JOSS)* to oversee the INSNA's relations with the publications, selection of *CONNECTIONS* and *JOSS*'s future editors and to coordinate the publications so that they are complimentary rather than in competition with one another. To insure openness to new ideas, one or more additional members will be selected by the President.

INSNA Awards 2010

- Freeman Award:
The 2010 Award went to Arnout van de Rijt, an Assistant professor of Sociology, at SUNY Stony Brook.
- Simmel Award:
The 2010 winner goes to Kathleen Carley who will accept the award as the keynote speaker at the 2011 Sunbelt Social Networks Conference in Florida.
- Student Award:
The 2010 Award went to Christophe Stadtfeld and Andreas Geyer-Schulz from the KIT Karlsruhe Institute of Technology for their paper "Analysing Event Stream Dynamics in Two Mode Networks."
- INSNA i2 Citation Award:
The 2010 winner of the INSNA citation award sponsored by i2 is Mark E.J. Newman from the University of Michigan, for his paper "A Measure of Betweenness Centrality Based on Random Walks, *Social Networks*, 27(1), 39-54, 2005."

CONNECTIONS

Manuscripts selected for publication are done so based on a peer-review process. See instructions to authors for information on manuscript submission. The journal is edited and published by the Connections Editorial Group:

Dr. Thomas Valente, Editor

Professor, Director of the Master of Public Health Program
Professor in the Department of Preventive Medicine,
University of Southern California, Alhambra, CA 91803

Dr. Kathryn Coronges, Managing Editor

Assistant Professor, Department of Behavioral Sciences & Leadership
United States Military Academy, West Point, NY 10996

Joseph Dunn, Associate Editor

305 Route 403, Garrison, New York 10524

Editorial Headquarters

University of Southern California, Institute of Prevention Research
1000 Fremont Ave., Unit #8, Building A, Room 5133, Alhambra, CA 91803
Tel: (626) 457-4139; fax: (626) 457-6699

Email tvalente@usc.edu or kate.coronges@usma.edu for questions or change in address. Published articles are protected by both United States Copyright Law and International Treaty provisions. All rights are reserved (ISSN 0226-1776).

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\$60 (\$40 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 items that lend themselves to electronic storage.

Sunbelt Social Network Conferences

Annual conferences for INSNA members take place in the United States for two years and in Europe every third year. The Sunbelt Conferences bring researchers together from all over the world to share current theoretical, empirical and methodological findings around social networks. Information on the annual Sunbelt Social Network Conferences can also be found on the INSNA website. Sunbelt XXXI will be held in St. Pete Beach, Florida from February 8 - 15, 2011.

Letters to the Editor

As a new feature of CONNECTIONS, we welcome letters to the Editor on all issues of academic interest. Letters should be brief and will be subject to editing and condensation. In addition, book reviews, network images or review articles will be considered for publication.

Instructions to Authors

CONNECTIONS publishes original empirical, theoretical, tutorial, and methodological articles that use social network analysis. The journal publishes significant work from any domain that is relevant to social network applications and methods. Commentaries or short papers in response to previous articles published in the journal are considered for publication. Review articles that critically synthesize a body of published research are also considered, but normally are included by invitation only. Authors who wish to submit a commentary, book review, network image or review article are welcome to do so.

Submitting Manuscripts

Authors are required to submit manuscripts online to the editor, Thomas Valente at *tvalente@usc.edu*. Expect a notice of receipt of your manuscript via email within one week. Feedback from the editor and reviewers will be sent to the corresponding author within six months after receipt. Revised or resubmitted manuscripts should include a detailed explanation of how the author has dealt with each of the reviewer's and editor's comments. For questions or concerns about the submission process, authors should contact the editor.

Manuscripts must be in MS Word format and should not exceed 40 pages including tables, figures and references. Manuscripts should be arranged in the following order: title page, abstract, corresponding author contact information, acknowledgments, text, references, and appendices. Abstracts should be limited to 250 words. Please embed all images, tables and figures in the document. Format and style of manuscript and references should conform to the conventions specified in the latest edition of Publication Manual of the *American Psychological Association*. Please disable any automatic formatting when possible. A figure and its legend should be sufficiently informative that the results can be understood without reference to the text. Every issue, we select an image from an accepted article to appear on the front cover of the journal.

CONNECTIONS

December, 2010

Volume 30 • Issue 2

ARTICLES

- Structural Redundancy and Multiplicity in Corporate Networks**..... 4
Roy Barnes and Tracy Burkett
- A Configurational Approach to Homophily Using Lattice Visualization**..... 21
David R. Schaefer
- Egocentric Networks and Unique and Shared Sources of Social Support**..... 41
Jeffrey A. Hall
- What You Believe Travels Differently: Information and Infection Dynamics
Across Sub-Networks**..... 50
Patrick Grim, Christopher Reade, Daniel J. Singer, Steven Fisher, and Stephen Majewicz

Structural Redundancy and Multiplicity in Corporate Networks

Roy Barnes

*Department of Sociology, Anthropology & Criminal Justice
University of Michigan-Flint*

Tracy Burkett

*Department of Sociology & Anthropology
College of Charleston*

Abstract

This research presents an intuitive and straight forward method of capturing both structural redundancy and the multiplicity of social ties in a small network of 20 corporate directors across four different social spheres in 1962. Structural redundancy is best thought of as the opposite of a unique tie which emerges in a network of interlocking corporate directors once the affiliations from other non-corporate organizations are included. Unlike the analysis of structural redundancy, the multiplicity of ties recognizes that the number and the configuration of ties between a given pair of directors are both meaningful. By utilizing these concepts, the paper shows how social club ties in 1962 were especially important in adding unique (i.e., non-redundant) ties among the corporate directors. The analysis of multiplicity reveals that over 60 percent of the directors had multiple ties and that 56 percent of the directors possessed ties stemming from two or more different types of social affiliations. These results underscore that there is more to the social cohesion among corporate directors than interlocks alone.

Authors

Roy Barnes has a Ph.D. in Sociology from the University of Wisconsin-Madison, and is an Associate Professor of Sociology and serves as the Assistant Dean of the College of Arts and Sciences at the University of Michigan-Flint. His research interests are in power structure research, interlocking directorates and network analyses.

Tracy Burkett has a Ph.D. in Sociology from the University of South Carolina, and is an Associate Professor of Sociology and Director of the Undergraduate Environmental Studies Program at the College of Charleston. Her research interests include political networks, sustainable community structures, and environmental justice.

Notes

An earlier version of this paper was presented at the Annual Congress of Humanities and Social Sciences, Society for Socialist Studies, 3 June 2010, The Political Economy of Corporate Power. The authors would like to thank Wouter de Nooy and Eelke Hemskerk for their helpful comments on an earlier version of this paper. We also remain grateful for the constructive comments provided by *Connections* in response to our initial submission.

Please address correspondence to Dr Roy Barnes, The University of Michigan-Flint, 522 French Hall , Flint, Michigan 48502; Email: rcbarnes@umflint.edu

INTRODUCTION

At the second meeting of the Politics and Interlocking Directorates Research Community, Burris (2006, p. 2) called for a “Sociology of Elites” in which corporate interlocks would be “only one thread in a much denser fabric of social ties among corporations and corporate elites.” Moreover, Burris cautioned those interested in pursuing a research agenda involving interlocking directorates “not to reify or isolate director interlocks from other social networks within which they are embedded.” This paper is a methodological response to this charge and explores the concepts of structural redundancy and multiplicity.

There is certainly a rich literature on the sociology of elites. The early works of Mills (1956) and Domhoff (1967, 1970, 1983, 1998) which document and map the inter-organizational ties among the directors of the United States’ major corporations are clearly foundational to contemporary analyses of the structure and exercise of economic and political power in America. For example, Thomas Dye’s work (2001) on his so-called “top down policymaking” begins with the national, i.e., corporate, elite. Similarly, Useem’s concept of the “Inner Circle” emphasizes how interlocking corporate directors obtain information on “contemporary business practices and the general business environment” (1984, p. 85). Likewise, in their study of policy domains, Laumann and Knoke (1987) envision dyadic communication among decision-makers as a means of monitoring streams of activities that may present opportunities or threats to their organizations. However, the men and women of the corporate elite are simultaneously directors of corporations and trustees or directors of museums, research universities, and members of social clubs. The notion that elites fulfill multiple roles and positions within a number of different types of organizations is captured by Breiger’s concept of “duality” (1974). Subsequently, many researchers, including those mentioned above, have examined ties between corporations and a variety of non-corporate organizations as a means of assessing the

interpersonal network of the business community.

In 1979, Moore published an article on the national elite network using data from the American Leadership Study from 1971-72. Based on personal interviews and the snowball sampling of other leaders, Moore constructed an interpersonal network that was used to identify a “central circle” of influential elites. Given this network clique, Moore was able to determine which individual business leaders were members of this “central circle” and whether such membership was associated with higher participation in non-profit foundations and policy planning organizations. In the early 1980s, Domhoff and a series of collaborators (Bonacich & Domhoff 1981; Salzman & Domhoff 1980, 1983) produced an extremely interesting set of findings concerning the links between corporations and policy planning groups, social clubs, universities, foundations and cultural/civic organizations.

While certainly praiseworthy for establishing the empirical foundations for the study of the inter-organizational ties between the corporate elite and policy planning organizations, nonprofit foundations and cultural organizations, these studies are not without problems. To begin, Useem’s and Domhoff’s analyses share the difficulty of operating solely at the organizational level of analysis. As such, in addition to analyzing inter-organizational connections, research into the formation, structure, and operation of the social mosaic of corporate directors would benefit from examining the inter-personal ties. Although Moore’s work operates at this interpersonal level of analysis, the reputational and snowball sampling methodology do not allow the complete analysis of the interpersonal network formed by corporate and noncorporate directors. These criticisms do not apply to the work by Carroll, Fennema and Heemskerk (2010) which examines directors among the global 500 corporations and the European Round Table of Industrialists, or to Carroll’s earlier work which examined the links between corporate directors and five leading transnational policy groups (2004). However, by focusing on one type of

extra-corporate relation, the issue of capturing the multiplicity of ties does not arise in these studies.

One of the classic problems accompanying the examination of affiliation network data is the loss of information when relations and ties are reduced from two-modes to a single mode (Field, Frank, Schiller, Riegle-Crumb & Mulle, 2006). When networks are reduced to a single mode, the duality of social experience – the intersection of actors and events described by Breiger (1974) – is no longer fully captured. As noted by Bonacich and Domhoff (1981), there are few methods that preserve the duality of persons and groups. Instead, researchers must rely on techniques that focus on a single set of actors or events, such as the directors of large American corporations in the present study. The methods that do exist to preserve information from both modes have focused chiefly on identifying sets of actors, clusters, and structural positions, rather than the contributions and configurations of types of ties across social space (See Laumann & Knoke 1987, Skvoretz & Faust 1999 and Doreian et al 2004 for examples of techniques.)

A second problem with affiliation data (as with all valued networks) is the loss of information that occurs when valued ties are dichotomized in order to facilitate analysis. As noted by Thomas and Blitzstein (2009), the transformation of valued data to binary data creates analytical uncertainty and may result in the loss of significant information. In the case of an affiliation matrix of corporate directors, the value of each tie represents the number of joint corporate, museum, university and social club affiliations shared by a set of directors. When ties are simply dichotomized, the relative value or the importance of any given type of relation is obscured. Opsahl, Agneessens and Skvoretz (2010) have recently introduced a refined set of algorithms to measure mode centrality in weighted (valued) networks. While certainly a step forward, these techniques still do not preserve the information captured by the multiplicity of ties. That is, they still do not capture which relations are the most important to the actors and the network.

In this paper, we describe a method of preserving important information about the context and configuration of ties while still allowing for ease of analysis with standard algorithms such as measures of centrality and density. Our method is unique in that it allows us to understand the context in which ties are formed and to judge the types of relational ties that are most important or valuable. Central to this project of answering Burris' call for a richer sociology of elites are the concepts of structural redundancy and the multiplicity of ties.

Structural Redundancy and the Multiplicity of Ties

Redundancy

Structural redundancy is best thought of as the opposite of a unique effect which occurs when new ties emerge in a network of interlocking corporate directors once the affiliations from other non-corporate organizations are included. On a more technical note, if additional ties are detected following the addition of two (or more) affiliation matrices and dichotomization of those ties, then unique ties exist. Similarly, if there are differences in the structure of the network generated by the affiliations from two (or more) organizations in comparison to the network of corporate ties alone, then the ties added are not redundant to the existing ties. In this paper, we will use four simple measures of network structure – network density, average degree, average betweenness, and average geodesic distance to highlight the unique contribution made by additional ties, or highlighting which ties are not redundant.

Methodologically, capturing the unique effects of adding one set of ties is straightforward and consists of simply comparing the network density, average degree, average betweenness and average geodesic distance of a network of directors without and with the additional ties. However, the methods are somewhat more complicated if we want to consider the unique effects of adding, say, the ties from university board memberships while controlling for the effects of all the other non-corporate ties. Here, we would need to compare a different set of

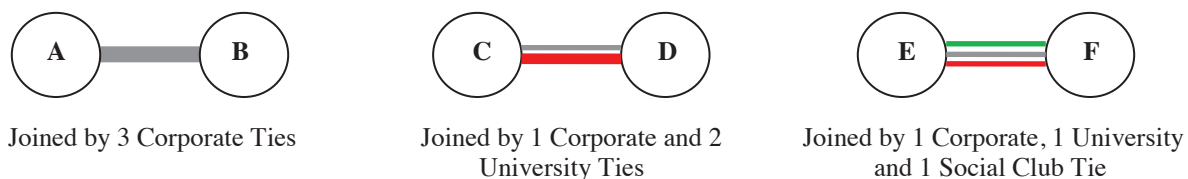
results – the measures of network structure first utilizing the ties formed from *all* of the various affiliations with the same measures calculated *without* the organization in question. In other words, if the ties formed through shared university board memberships are redundant, then there should be no difference between the measures based on all ties and those same measures based on all ties while excluding those from university boards. Alternatively stated, if the analysis of the network with all but the university-based ties yields lower average degree or betweenness centrality, and higher average geodesic distance, then we can conclude that university ties have indeed made a unique contribution to the structure of the network.

The second set of methodological issues concerns capturing the multiplicity of ties. Unlike the analysis of structural redundancy where we intentionally ignore multiple ties (i.e., we dichotomize the ties between a given pair of directors so that the tie is present or not), the multiplicity of ties recognizes that the number of ties between a given pair of directors matters. Furthermore, not only is the number of ties

important, the *configuration* of those ties is also meaningful. Consider Figure 1 which depicts three pairs of directors – all of which possess three ties between them.

As seen in Figure 1, there are three types of ties represented – a tie formed by serving on the same corporate board, ties formed by being on the same university board, and a tie generated by belonging to the same social club. Furthermore, the relative thickness of the line indicates the number of ties. By extending the multiplicity of ties beyond the number of ties, this approach allows a researcher to identify those directors possessing only corporate ties, versus those associated by both corporate and university ties, versus those directors that are linked through corporate and two or more other organizational ties as seen in the last example. By capturing the configuration of ties, this approach will be able to distinguish if a given pair of directors met only in boardrooms, or whether there were opportunities to meet outside board of director meetings – at a monthly Regents’ meeting or on the golf course during the weekend.

Figure 1. Disaggregating the Multiplicity of Ties Among Corporate Directors



DATA

To more clearly illustrate these concepts and our methodology, let us consider the following set of directors. These affiliations represent actual ties abstracted from the much larger study of corporate directors from the year 1962.¹ At its

core, the network consists of 11 of the 23 directors that served on the board of the First National Bank of Chicago. The remaining 9 directors in this example had affiliations with 14 other corporations. Although not representing all of the types of ties in this study, some of these 20 directors also served on 2 museum boards, 2 university boards, and self-reported membership in 5 social clubs. Taken together,

¹ All of the data have been obtained from public sources. The corporate boards of directors were taken from various *Moody's* manuals. The memberships on museum boards and university boards were gathered from annual reports or from

university documents. Data on social club membership were obtained from the biographical sketches published in *Who's Who in America*.

Table 1. 2-Mode networks of 20 Directors by 24 Organizational Affiliations

Director	Corporations														Museum		Univ.		Social Clubs					
	Armour & Co.	Borg-Warner Corp.	Caterpillar Tractor Co.	Chase Manhattan Bank	Commonwealth Edison Co.	Container Corp of America	Continental Ill. Nat. Bank & Trust	Equitable Life Assurance	First National Bank of Chicago	Inland Steel	International Harvester	John Hancock Mutual	Sears, Roebuck & Co.	Standard Oil (Ind)	Swift & Co.	Art Institute of Chicago	Museum of Science & Industry	Northwestern university	University of Chicago	Century	Chicago	Commercial	Indian Hill	Links
Barr John A.					x													x		x				
Block Joseph L.					x				x									x			x			
Chewning E. Taylor	x										x													
Clark P. F.	x										x										x			
Cushman A. T.									x			x												
Eberhard H. S.			x						x															
Freeman G. A. Jr.		x							x															
Gale Willis										x														
Goodrich Paul W.													x											
Ingersoll R. S.		x							x															
Jarvis P. M.																								
Kennedy David M.																								
Livingston H. J.																								
McCormick Brooks																								
McDowell Remick																								
Oates James F. Jr.																								
Prince W. W.																								
Rockefeller David																								
Swearingen John E.																								
Ward J. Harris																								
Total Affiliations	4	2	2	3	5	2	3	2	11	3	5	3	2	2	6	2	4	3	5	2	12	12	2	2

Table 1 presents the affiliations of these 20 directors across the 24 different organizations. When transformed into separate 20 x 20 affiliation matrices representing corporate ties only, museum board ties only, university board ties only and social club membership ties only, each section of the 2-mode network generated a distinct set of affiliations. Based on corporate ties alone, there were 90 unique ties among the directors. There were 7 ties formed between directors through shared affiliations while serving on the two museum boards. Service on the boards of Regents or Trustees of the two universities created 13 ties. Finally, in 1962, these 20 directors had 97 ties through shared social club memberships.

Considering first the question of structural redundancy, the individual images in Figure 2 below show in dramatic fashion how the addition of non-corporate ties (museum ties, ties based on university board service, and affiliations through shared social club memberships) represent unique contributions to the network's structure.²

While these visualizations are useful for networks of 20 directors, summary measures of network structure are needed to precisely capture the unique effects of ties from non-corporate organizations and to accommodate larger networks. Table 2 lays the foundation of these measures by reporting the number of total possible ties, the number of unique ties, and redundant ties. The first column is simply the number of possible ties in a network of 20 directors.

² To add the affiliations from the different organizations, there are two options within UCINET. Begin by creating four 2-mode networks – the 20 directors by each set of organizational ties. The first option is to join the columns of the 2-mode matrices of interest. After this command is executed, one can use the 2-mode to 1-mode command to generate the final affiliation matrix. Alternatively, one can convert each of the four 2-mode networks into their respective affiliation matrices. Then, to combine the sets of affiliations of interest, you use the matrix algebra tool in UCINET to add the matrices. Once you have joined the matrices and have your affiliation matrix, you can dichotomize the ties.

To determine the number of unique ties, the analyses utilize binary ties between directors. Consequently, the addition of a tie through shared board memberships on the Art Institute of Chicago will not add to the network if the tie already existed by virtue of being on the same corporate board of directors. Therefore, if one compares the 90 ties that are formed through shared corporate affiliations with the 93 unique ties that emerge when one combines the corporate and university based ties, one can then conclude that of the 7 ties formed from the museum ties alone, 4 of them were redundant. In contrast to these relatively small numbers, consider the 133 unique ties that emerge when we combine corporate and club organizations. Since there were 97 ties among these directors based solely on social club memberships, we can then conclude that from these 5 private social clubs, 43 unique ties emerged while 54 were redundant to ties already created by shared corporate affiliations.

This discussion of the number of unique ties leads seamlessly to the first measure of network structure, network density. Network density is simply the number of unique ties divided by the number of all possible ties. Since there are 90 ties formed through shared corporate affiliations, and there are 190 total possible ties in a network of 20 directors, the density = $90/190 = 0.4737$. When the 4 unique ties from museum ties are added, the network density increases slightly to 0.4895. In contrast, when the non-redundant ties from social clubs are added to the network of interlocking directorates, the network density increases to 0.7000. The second half of Table 2 compares the network density based on all four types of social affiliations with the density when a specific type of organization is excluded. Based on all of the ties, the network density was 0.7211. When the ties based on museum affiliations are excluded, the network density remains unchanged. Therefore, museum ties did not contribute any unique ties when the social ties generated from corporate, university and social clubs are considered. In contrast, when we compare the network density from all sources with the density excluding university ties, we find a slight decrease in network density. The exclusion of ties from social clubs

Figure 2. Unique Contribution of Non-Corporate Ties

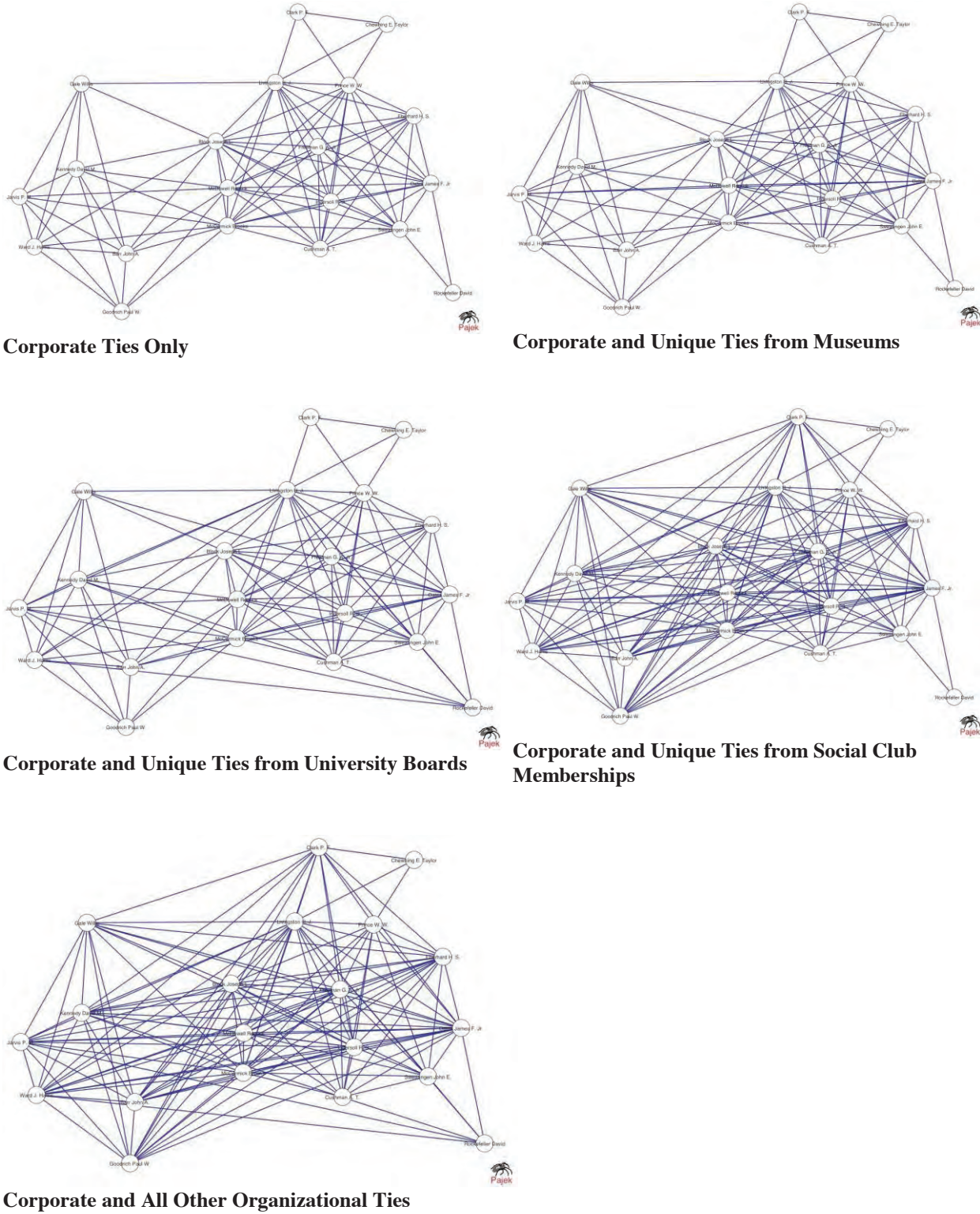


Table 2. Changes in Network Ties with the Addition and Exclusion of Non-Corporate Ties

	N of Possible	N of Unique	N of Redundant	N of Unique Ties From Excluded Org	Network Density
Corporation Ties Alone	190	90			0.4737
Corporation and Museum	190	93	4		0.4895
Corporation and University	190	99	4		0.5211
Corporation and Social Clubs	190	133	54		0.7000
Corporation and All Other Organizations	190	137	70		0.7211
All Ties Excluding Museum	190	137	7	0	0.7211
All Ties Excluding University	190	133	9	4	0.7000
All Ties Excluding Social Club	190	101	54	36	0.5316

Table 3. Changes in Network Structure with the Addition and Exclusion of Non-Corporate Ties

	Average Degree		Average Betweenness		Average Geodesic Distance	
Corporation Ties Alone	9.0		0.0345		1.62	
Corporation and Museum	9.3	†	0.0330	ns	1.59	*
Corporation and University	9.9	*	0.0278	ns	1.50	**
Corporation and Social Clubs	13.3	***	0.0170	ns	1.31	***
Corporation and All Other Organizations	13.7		0.0155	na	1.28	
All Ties Excluding Museum	13.7	na	0.0155	ns	1.28	na
All Ties Excluding University	13.3	*	0.0170	ns	1.31	*
All Ties Excluding Social Club	10.1	**	0.0272		1.49	***

^{na} Wilcoxon Signed Ranks Test is not applicable since there was absolutely no difference between means

^{ns} Wilcoxon Signed Ranks Test did not find a difference between ranks among directors

† Wilcoxon Signed Ranks Test of the difference between ranks among directors significant at the 0.10 level.

* Wilcoxon Signed Ranks Test of the difference between ranks among directors significant at the 0.05 level

** Wilcoxon Signed Ranks Test of the difference between ranks among directors significant at the 0.01 level

*** Wilcoxon Signed Ranks Test of the difference between ranks among directors significant at the 0.001 level

has a dramatic effect on overall network density – from 0.7211 to 0.5316.

In addition to overall network density, we can also capture the effects of these unique ties from museums, research universities, and membership in elite social clubs on the network of corporate ties by comparing average measures of individual centrality and cohesion within the network. The two measures of centrality – degree and betweenness – and the single measure of cohesion – geodesic distance – require a more detailed introduction. Degree centrality is simply the number of other directors directly tied to a given director (Freeman, 1979). Betweenness centrality is the average of the proportions in which a given director lies on the shortest path between all other pairs of directors (Freeman, 1979). While this is a more complicated measure, it essentially captures how important a given director is in serving as a link between other directors. The final measure of network structure we employ is that of average geodesic distance. The geodesic distance is the shortest path between two directors. While not a measure of network importance, geodesic distance is a measure of cohesion and as such, a director that has low average geodesic distance will be closer, on average, to the other directors in the network.

Having defined the summary measures of network structure, we can now turn to Table 3 that reports the average degree, average betweenness, and average geodesic distance for the small network of 20 directors. Given that the distributions of these variables are not normal, Table 3 also reports the significance levels from the Wilcoxon Signed Ranks tests.

The top half of the table shows how adding affiliations from university boards and social clubs significantly changes average degree and geodesic distance, but not betweenness centrality. There is a significant increase in average degree among these 20 directors when adding the unique ties from university boards and a very significant increase in average degree when adding non-redundant social club ties. In contrast, the results for betweenness centrality do not show a significant difference and in fact,

trend in a counterintuitive way. For example, when ties from universities are added the average betweenness centrality is noticeably lower than for the network based on corporate ties alone. While it might be expected that as the number of unique ties increases the centrality of the directors would increase, in adding new ties, we actually increase the number of direct ties and therefore reduce the number of instances where a given director will lie on the shortest geodesic between two other directors. Finally, regarding average geodesic distance, there were significant network changes when the ties from each of the three non-corporate organizations are added. While there is only a 0.03 decrease in average geodesic distance when museum ties are added, this nonetheless represents a significant decrease. Substantively larger decreases in geodesic distance are seen when university and social club ties are included. Indeed, when social club ties alone are added, the social distance drops from 1.62 steps to 1.31 steps.

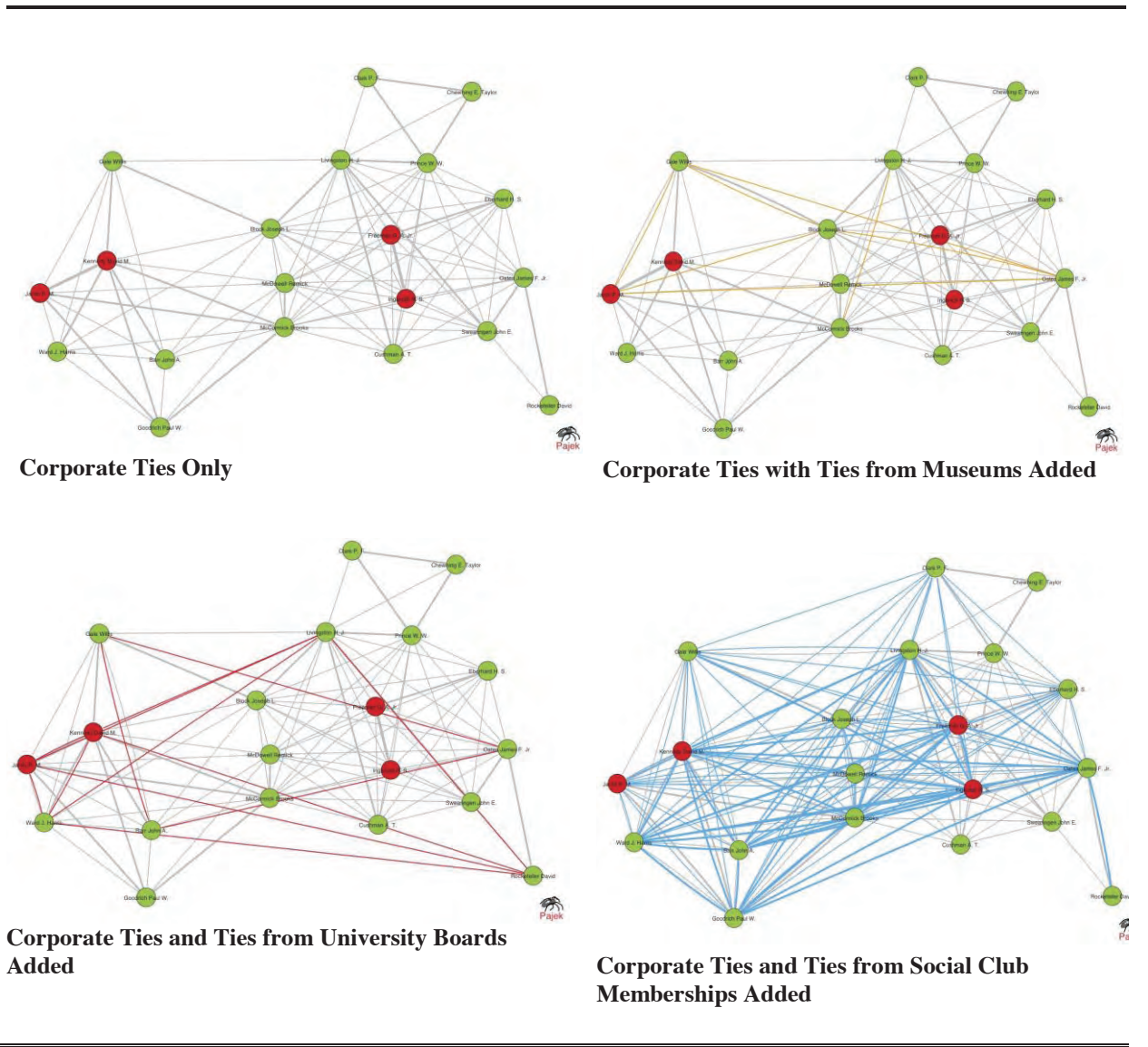
The bottom half of Table 3 presents the results when one compares average degree, betweenness, and average geodesic distance based on a network of all ties, with the network of ties excluding the organization of interest. While adding university ties to corporate ties may generate unique ties that increase average degree or decrease geodesic distance, it is not necessarily the case that when we consider all sources of social ties, universities will still add unique ties not previously in existence from museum boards or social clubs. As it turns out, museums do not add to the average degree of directors nor lower the average geodesic distance. However, when we compare the average degree and geodesic distance based on all ties with these measures based on all ties except university ties, we do find significant differences in average degree centrality and geodesic distance. Even more dramatic differences emerge when we compare the average degree and geodesic distance based on all affiliations versus these measures excluding the ties from social clubs. Average degree centrality falls from 13.7 to 10.1 and the average social distance increases from 1.28 to 1.49.

Multiplicity

While this brief example has documented the unique contributions of a variety of social ties, it also suggests that only corporations and clubs are significant in forming the structure of the network. In terms of structural redundancy, this is certainly true. However, by treating all ties as binary, the above analyses discard rich information on the multiplicity of social ties that exist among these 20 corporate directors. Therefore, to complement the above analyses of *structural redundancy*, the following analyses next consider the *multiplicity of ties*. We have

chosen to use the term multiplicity in contrast to multiplexity since we are not solely concerned with modeling networks based on multiple ties. The goal of our method is to capture the *configuration* of social ties, not simply their presence or absence or even the simple number of ties. Therefore, we wish to answer the following types of questions. Are directors tied exclusively or primarily through corporate ties? Do ties from other sources of social interaction complement or even outweigh the ties created through overlapping memberships on corporate boards?

Figure 3. Example of 20 Directors and Sources of Social Ties



Beginning with the set of visualizations in Figure 3, the analyses use the network of corporate ties as a point of departure. Note that the widths of the arcs vary to represent the strength of the particular tie and the colors of the vertices are coded to represent their highest level of multiplicity. For example, in the network of corporate ties alone, we see two pairs of directors that are red indicating that each has at least one tie that is of strength 3 – meaning that there are three corporate ties linking these two directors. Further, the tie between the two red vertices is clearly thicker than the other edges in the network which have strengths of 1 or 2. Finally, all of the other vertices in the network are green indicating that they have at least one tie with a strength value of 2.

In the subsequent networks in Figure 3, ties from each set of noncorporate organization are superimposed on the network of corporate ties. This is accomplished by treating corporate ties as arcs in one direction while the ties from say museums are graphed as arcs in the opposite direction. Looking at the network of corporate ties with museum ties added, we see first, that there are relatively few ties added. Furthermore, if we examine the yellow orange ties that are

added from shared affiliations with museums, the majority of the arcs are bi-directed indicating that they simply reinforce an existing corporate tie. In contrast, note the third graph in which the red ties in the lower right corner of the graph are clearly new ties formed through university board affiliations (although the tie between Rockefeller and Livingston is slightly obscured by other corporate ties). Finally, the fourth graph with cyan ties dramatically underscores the number of ties between these 20 directors that are formed through shared club memberships.

While the different configuration of each individual set of ties is certainly noteworthy, the sociographs in Figure 3 do not allow us to display the *simultaneous* set of ties from all organizations. To pull all of this information together and to capture multiple types of ties between a given pair of directors, it is necessary to aggregate the ties so as to preserve the types and strengths of all of the ties. To accomplish this condensing/preservation of information, we have assigned two values to each tie between directors – a strength code and a type code. Consider the following sample of seven ties between John Barr and three other directors (Table 4).

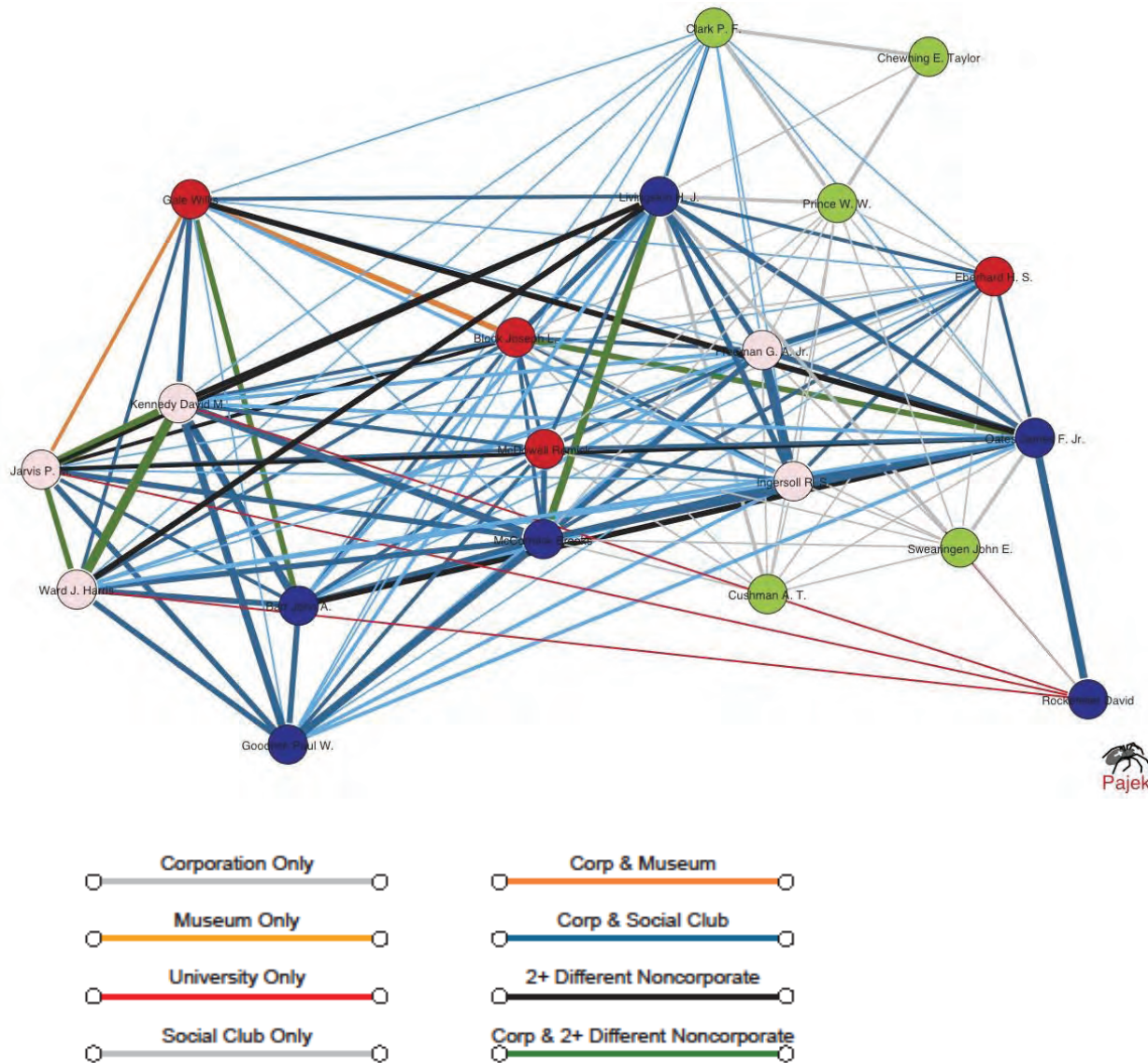
Table 4. Individual Ties with John A. Barr (01)

Tie ID	Director (vertex #)	Strength	Type Code	Type of Tie
1.02	Block, Joseph L. (02)	1	100000	Corporation
1.02	Block, Joseph L. (02)	<u>1</u>	<u>1</u>	Club
		2	100001	Aggregated Sum
1.16	Oates, James F. Jr. (16)	1	100	University
1.16	Oates, James F. Jr. (16)	<u>2</u>	<u>1</u>	Club
		3	101	Aggregated Sum
1.08	Gale, Willis (08)	1	100000	Corporation
1.08	Gale, Willis (08)	1	100	University
1.08	Gale, Willis (08)	<u>1</u>	<u>1</u>	Club
		3	100101	Aggregated Sum

There are two ties (coded with an ID of 1.02) between John A. Barr and Joseph Block – one based on being on the same board of directors and one due to mutual membership in a social club. Once we aggregate these two ties on the basis of the unique tie ID of 1.02 and sum the strength and type codes, the result is that there is now one tie between John Barr and Joseph Block that has a strength value of 2 and a configuration of ties code equal to 100001. This code for the tie between these two directors can

now be directly interpreted as a tie consisting of two affiliations, one from a corporate board and the other from a social club. For the second set of director ties with an ID of 1.16, John Barr has two different types of ties with James F. Oats Jr. Unlike the first example however, the tie based on club memberships has a strength value of 2. When aggregated, the resulting tie between Barr and Oats represents three ties– one university based and two club based ties. The aggregated code for the types of ties between Barr and Oats

Figure 4. Corporate Ties With All Other Organizational Ties Added



is 101 meaning that these two directors are tied together through both university and club based ties. Finally, turning to ties with an ID of 1.08, we find three types of ties between John Barr and Willis Gale, each tie with a strength value of 1. When we aggregate these ties, summing the strength and type codes, we obtain a tie with a strength value of 3 and a code of 100101 meaning that these directors are tied through corporate, university and club ties.

Given the results of aggregating each of the individual types of ties between all of the directors, it also becomes necessary to distinguish each of the four different types of ties. Therefore, corporate ties are gray, ties from museums are yellow orange, ties from universities are red and social club ties are cyan.

However, we also need to incorporate four new colors – ties that consist of shared corporate and museums are orange; ties between directors that are based in both shared corporate board and club memberships are midnight blue; ties that are formed from two or more non- corporate organizations are black; and finally, ties that are based in corporate ties and at least two other non-corporate affiliation are olive green. If we consider the ties in Table 4, the set of ties between John Barr and Joseph Black would appear as a single midnight blue edge with a strength value of 2; the ties between John Barr and James Oats would now be a black edge with a strength value of 3; and the three individual ties between John Barr and Willis Gale would now be represented by a single olive green edge with a strength value of 3. See Figure 4.

Table 5. Descriptive Statistics on Ties Between Directors

	N	Percent
Distribution of Tie Strength		
Single Tie Between Directors	53	38.69
Two Ties between directors	48	35.04
Three Ties between directors	27	19.71
Four Ties between directors	6	4.38
Five Ties between directors	3	2.19
Total	137	100.01
Number of the Different Types of Ties between Directors		
Only One Type of Tie	74	54.01
Two Different Types of Ties	56	40.88
Three Different Types of Ties	7	5.11
Total	137	100.00
Distribution of the Types of Ties		
Corporate Tie Only	34	24.82
University Tie Only	4	2.92
Social Club Tie Only	36	26.28
<i>Subtotal</i>	74	54.02
Corporate and Museum Ties	2	1.46
Corporate and Social Club Ties	48	35.04
<i>Subtotal</i>	50	36.50
Two or more Types of Non-Corporate Ties	7	5.11
Corporate and Two or more Non-Corporate Ties	6	4.38
<i>Total</i>	137	100.01

On a final introductory note, our last visualization which represents the multiple types of ties between directors has new colors for the vertices. When all sources of social ties are considered simultaneously, some directors have at least one tie with another director with a strength value of 4 – for example, John Barr, Paul Goodrich and Brooks McCormick all have 4 ties between themselves. R.S. Ingersoll and G.A. Freeman Jr. (the pink vertices) each have 5 ties between themselves. The results are presented in our final graph, Figure 4 above, which pulls everything together and shows corporate ties and ties from all other organizations.

As with the sociographs of network structure, the graphs depicting the multiplicity of ties provide a clear overview of the constellation of the various ties constituting the social mosaic among these 20 corporate directors. They are unfortunately equally ill-suited to provide clear measurements of the configuration of types of ties and the multiplicity of those ties. However, by aggregating the individual types of ties between a given pair of directors, we can easily obtain three summary measures of the 137 ties between directors – the distribution of ties by strength, the distribution of the number of types of ties, and finally, the distribution of types of ties themselves.

Table 6. Configuration of Ties Involving John A. Barr

Vertex	Director	Strength	Type	Proportion
2	Block Joseph L.	1	Corporation	} 9/33= 0.2727
8	Gale Willis	1	Corporation	
9	Goodrich Paul W.	1	Corporation	
11	Jarvis P. M.	1	Corporation	
12	Kennedy David M.	2	Corporation	
14	McCormick Brooks	1	Corporation	
15	McDowell Remick	1	Corporation	
20	Prince W. W.	1	Corporation	
8	Gale Willis	1	University	
16	Oates James F. Jr.	1	University	
2	Block Joseph L.	1	Club	} 22/33= 0.6667 Predominantly Club Based
4	Clark P. F.	1	Club	
6	Eberhard H. S.	1	Club	
7	Freeman G. A. Jr.	2	Club	
8	Gale Willis	1	Club	
9	Goodrich Paul W.	2	Club	
10	Ingersoll R. S.	2	Club	
11	Jarvis P. M.	1	Club	
12	Kennedy David M.	2	Club	
13	Livingston H. J.	2	Club	
14	McCormick Brooks	2	Club	
15	McDowell Remick	1	Club	
16	Oates James F. Jr.	2	Club	
20	Ward J. Harris	2	Club	
Total Ties		33		

The last section of Table 5 summarizes the precise configurations of ties. Recall in our discussion of Figure 4, there were potentially eight types of ties. However, when aggregated, only seven distinct patterns of social affiliations are represented—corporate only, university only and social club only (note there were no ties that were only based on shared affiliations through museums since all of those ties were structurally redundant). Table 5 shows that only a quarter of the ties were from corporate ties alone. While fewer than 3 percent of the ties were formed solely through shared university ties, over a quarter (26.28 percent) of the ties were based exclusively on mutual social club memberships.

The bottom section of Table 5 also shows that over a third of the ties between these 20 directors were constituted by both corporate and social club ties. Equally important from the perspective of emphasizing the role of non-corporate ties in integrating our corporate directors, 7 ties were based on two or more noncorporate ties. Finally, Table 5 indicates that 6 of the 137 ties were from three or more sources, including at least one corporate tie. In short, the results in Table 5 underscore the message from Figure 4 – non-corporate ties constitute a substantial portion of the social connections between our example directors, with social club ties being most abundant.

Thus far, the discussion has focused on how to represent the multiplicity of ties between the directors in this example network. However, it would also be useful to further aggregate these various configurations of ties back to the individual directors. To accomplish this aggregation, the analyses revert to the original set of 207 ties. However, rather than aggregating on the basis of the tie ID itself (1.02, 1.16 or 1.08), this time we need to aggregate by the vertex number (making sure that both sending and receiving vertices are represented). Therefore, by shifting the level of analysis back to the directors themselves, each director will possess a set of ties that can be characterized by being completely or predominantly corporate. In contrast, a director’s set of ties can be primarily grounded in other organizational affiliations. Consider again John A. Barr. This

time however, Table 6 presents all of his 33 ties with the 19 other directors.

By sorting the ties by type, it is straightforward to calculate both the sum of strengths which reflects the 33 individual ties, and the sum of the strengths for each type of tie.³ Table 6 shows that of the 33 ties, 9 are based on shared corporate board memberships, 2 on mutual membership on university Boards of Regents or Trustees, and 22 on shared social club memberships. This means that 27.3 percent of John Barr’s ties were corporate, 6 percent were university and 66.7 percent were social club based. When this procedure is applied to the remaining 19 directors, we can then obtain summary measures of the configuration of ties possessed by each director. Table 7 presents the summary statistics for the 20 directors in this example.

Table 7. Configuration of Ties for Directors

Average Proportion of Corporate Ties	0.4990
Average Proportion of Museum Ties	0.0215
Average Proportion of University Ties	0.0540
Average Proportion of Social Club Ties	0.4255

Therefore, on average, the twenty directors are characterized by half of their ties being corporate based, and nearly another half are based on social club ties. Finally, Table 7 indicates that the average percent of ties based on museum ties is only 2.15 percent and the average percent of ties based on shared university ties is 5.4 percent. In other words, this example of 20 directors is tied together by both corporate and club ties in roughly equal proportions. While this conclusion is far from

³ Calculating the degree of each director on the basis of each of the four affiliation matrices would yield the number of different directors tied to John Barr and as a consequence, we would lose the multiple corporate ties between this director and David Kennedy, and the multiple club ties between John Barr and G.A. Freeman Jr., Paul W. Goodrich, R.S. Ingersoll, David Kennedy, H.J. Livingston, Brooks McCormick, James F. Oats Jr., and J. Harris Ward.

startling given the sociographs in Figure 4, these aggregate measures can prove useful if one wishes to relate these measures of a director's configuration of ties to non-network characteristics such as the director's level of political contributions.

SUMMARY & CONCLUSIONS

In this paper we have expanded the literature on the sociology of elites by introducing a method which preserves the context and richness of their multiple relationships. As shown, the shared affiliations through museums, universities, social clubs, and of course, shared corporate board memberships, play varying, but distinct, roles in binding together these corporate directors. Using this intuitive procedure, we simultaneously capture the multiplicity of ties and assess their unique impact on corporate elite networks. That is, we are able to discern which types of ties make the greatest contribution to the director's network – a network which is derived from several co-memberships on organizational boards and social clubs.

While we have applied this technique to gaining greater insight into the social fabric in which the corporate elite reside, it can certainly apply to other affiliation networks with multiple sets of ties. Our technique is novel in that it allows researchers to use standard network measures while bypassing two shortcomings that accompany the analysis of affiliation networks: the loss of contextual information that occurs when networks are reduced from two modes to a single mode and the loss of information that occurs when these valued networks are dichotomized for ease of analysis.

Although the data presented here are for an artificially small set of social ties, this method can be employed to analyze larger networks. In fact, the method is highly flexible as researchers who adopt it may opt to investigate entire networks, specific network neighborhoods, or sets of actors depending on the theoretical and substantive interests.

REFERENCES

- Bonacich, P.F. & Domhoff, G.W. (1981). Latent Classes and Group Membership. *Social Networks*, 3, 175-196.
- Breiger, R. L. (1974). The Duality of Persons and Groups. *Social Forces*. 53, 181-190.
- Burris, V. (2006). Interlocking Directorates and Political Mobilization among Corporations and Corporate Elites: Evidence from the U.S. Paper presented at the *Politics and Interlocking Directorates Conference*, University of Barcelona, September 28-29.
- Carroll, W. K. (2004). *Corporate Power in a Globalizing World: A Study in Elite Social Organization*. Don Mills: Oxford University Press.
- Carroll, W.K., Fennema, M. & Heemskerck, E.M. (2010). Constituting Corporate Europe: A Study of Elite Social Organization. Paper submitted to Antipode.
- Domhoff, G.W. (1967). *Who Rules America?* Englewood Cliffs: Prentice-Hall.
- Domhoff, G.W. (1970). *Higher Circles*. New York: Random House.
- Domhoff, G.W. (1983). *Who Rules America Now?* Englewood Cliffs: Prentice-Hall.
- Domhoff, G.W. (1998). *Who Rules America? Power and Politics in the Year 2000*. Mountain View, CA: Mayfield Publishing Company.
- Doreian, P., Batagelj, V. & Ferligoj, A. (1994). Partitioning Networks Based on Generalized Concepts of Equivalence. *Journal of Mathematical Sociology*, 19, 1, 1-27.
- Dye, T.R. (2001). *Top Down Policy Making*. New York: Chatham House Publishers.
- Field, S., Frank, K.A., Schiller, K., Riegle-Crumb, C., & Mulle, C. (2006). Identifying Social Contexts in Affiliation Networks: Preserving the Duality of People and Events. *Social Networks*, 28, 97-123.
- Freeman, L. C. (1979). Centrality in social networks: I. Conceptual clarification. *Social Networks*, 1, 585-592.
- Laumann, E.O. & Knoke, D. (1987). *The Organizational State: Social Choice in National Policy Domains*. Madison, WI: University of Wisconsin Press.
- Mills, C.W. (1956). *The Power Elite*. New York: Oxford University Press.

- Moore, G. (1979). The Structure of a National Elite Network. *American Sociological Review*, 44, 673-692.
- Opsahl, T., Agneessens, F. & Skvoretz, J. (2010). Node Centrality in Weighted Networks: Generalizing Degree and Shortest Path. *Social Networks*, 32 (3), 245-251.
- Salzman, H. & Domhoff, G.W. (1980). The Corporate Community and Government: Do They Interlock? In G.W. Domhoff (Ed.), *Power Structure Research* (pp. 227-54). Newberry Park, CA: Sage Publications.
- Skvoretz, J. & Faust, K. (1999). Logit Models for Affiliation Networks. *Sociological Methodology*, 29, 253-278.
- Thomas, A.C. & Blitzstein, J. (2009). The Threshold Problem: Uncertainties Due to the Dichotomization of Valued Ties. Conference Proceedings, UAB.
- Useem, M. (1984). *The Inner Circle*. New York: Oxford University Press USA.

A Configurational Approach to Homophily Using Lattice Visualization

David R. Schaefer

*School of Social and Family Dynamics
Arizona State University*

Abstract

This research approaches homophily as a multidimensional concept and uses combinatorial logic to investigate the patterns of homophily that exist in relations from different substructures. Particular focus is on which patterns of homophily (i.e., configurations) occur more often than expected by chance given the demographic distribution of the population. Lattices are introduced as an intuitive way to represent homophily multi-dimensionally and uncover patterns within and across relation types. Qualitative Comparative Analysis is used as a means to uncover the conjunctural forces that underlie homophily. These techniques are applied to General Social Survey data on discussion relations. Results demonstrate the low, but nonzero, tolerance for dissimilarity in relations and how patterns and levels of homophily vary across relation types. The majority of relations exhibited homophily on only three or four of the five dimensions measured. Race was a key dimension of homophily for all relation types (in fact, necessary for kin and group member relations), while kin relations also required religious homophily and friend relations almost always required age homophily.

Author

David R. Schaefer is Assistant Professor in the School of Social and Family Dynamics at Arizona State University. His current research draws upon social network and social exchange theories to investigate the processes behind network formation and change. Recent works have appeared in *Social Networks*, *American Sociological Review*, and *Social Psychology Quarterly*.

Direct all correspondence to David R. Schaefer, School of Social and Family Dynamics, Tempe, AZ 85287-3701; Email: david.schaefer@asu.edu.

INTRODUCTION

Folk wisdom tells us that “birds of a feather flock together” yet, at the same time “opposites attract.” The contradiction between these familiar adages points to a fundamental feature of relationships: their multidimensionality. The forces that attract and repel people operate in conjunction along several dimensions simultaneously. This is evident in Blau’s (1977) notion of multiform heterogeneity, which implies that the desire for homophily on some dimensions can make heterophily on other dimensions inevitable. Consequently, individuals within relationships are similar in many respects, while dissimilar in other aspects: Male and female birds flock together.

This study moves beyond the unidimensional treatment of homophily that dominates previous research by developing a multidimensional conceptualization. Homophily is characterized as a combination of similarity and dissimilarity across several dimensions, referred to as *configurations of homophily*. These configurations can be analyzed to better understand the interrelations between dimensions that structure interpersonal associations. Of particular interest are which configurations of homophily occur more often than expected by chance.

Introducing multidimensional homophily proceeds through two steps. First, I define relations as configurations of homophily with the graphical assistance of lattices. Just as scatterplots are valuable for examining correlational data, lattices provide a useful representation of configurational data. I present a lattice visualization technique that conveys both the distribution of dyads across homophily configurations and the logical relationship between configurations. This tool helps portray the relationships inherent to combinatorial data and can assist in the evaluation of such data.

Second, to understand which combinations of homophily are most salient across relations, I examine ego network data from the 1985

General Social Survey. I use Qualitative Comparative Analysis (QCA) to identify the combinations of similarity and dissimilarity that underlie the configurations of homophily in observed dyads. This process uncovers how dimensions of homophily interact to create more dyads than expected by chance.

Homophily Research

Social scientists have documented homophily as one of the most persistent features of human society (McPherson, Smith-Lovin, & Cook, 2001). Homophily is defined as the tendency for people to associate with others who are similar, rather than dissimilar, to themselves. Homophily is important because of its implications for the movement of ideas and people through society. While communication between similar individuals tends to be more effective (Rogers & Bhowmik, 1970), homogeneous networks also tend to constrain attitudes to be more “closed-minded” (Laumann, 1963) or conservative (Bienenstock, Bonacich, & Oliver, 1990). Beyond attitudes, the presence of homophily necessitates that any “cultural, behavioral, genetic, or material information that flows through networks will tend to be localized” (McPherson et al., 2001, p. 416). Homophily also has consequences for social mobility to the extent that positions of power and authority are correlated with ascribed characteristics of their occupants. For instance, because women are less likely to occupy powerful positions within organizations, such individuals must form strategic, heterophilous ties that are less efficient in order to gain access to valuable information and resources (Ibarra 1992). Thus, it is important to understand homophily as a means of ascertaining the types of social distance that are least often bridged.

Research has documented the persistence of homophily in groups and interpersonal relationships across a wide range of dimensions (for an overview see McPherson et al., 2001). Race and ethnic homophily is perhaps the strongest dimension in the United States and is observed across most forms of relations.

Spouses (Kalmijn, 1998), coworkers (Ibarra, 1995), and neighbors (Marsden, 1990) all tend to be of similar race/ethnicity. Relations also tend to exhibit strong age homophily (Feld, 1982; Marsden, 1990) with the main exception of parent/child relations (Fischer, 1982). Sex is an unusual dimension in that for many people spouses are an important relation yet spouses are traditionally opposite sex. Excluding spouses, relations tend to be homophilous on sex, though not as much as on race/ethnicity (Marsden, 1987; Mayhew, McPherson, Rotolo, & Smith-Lovin, 1995). Homophily along the religion dimension is fairly strong, almost as strong as race/ethnic homophily (Marsden, 1981, 1988). Education homophily has also been demonstrated (Louch, 2000) though the presence of kin in networks reduces its strength (Kalmijn, 1998). Finally, homophily of values, status, and appearance are all important criteria in selecting long-term partners (Buston & Emlen, 2003).

Baseline Homophily

When measuring homophily it is essential to ascertain how much perceived similarity is simply due to chance. McPherson and colleagues (2001) distinguished between two sources of homophily: *baseline* and *inbreeding*. Baseline homophily is the level that would be expected by chance if people chose their associates randomly. Baseline homophily is purely a function of the population distribution: intergroup contact is more frequent when the population is more heterogeneous (Blau & Schwartz, 1984; Blum, 1985; Skvoretz, 1990). As the homogeneity of a population increases, relations within the population are expected to be more homophilous.

Inbreeding Homophily

When homophily on any dimension exceeds the baseline rate then that dimension is defined as *salient*; the number of relations that are ingroup on that dimension exceeds that expected by chance (Blau, 1977). Rates of homophily exceeding the baseline level are products of one or more inbreeding sources. Three primary

mechanisms contribute to inbreeding homophily: *consolidation*, *substructures*, and *preference* (McPherson et al., 2001). Consolidation emerges because individuals occupy social positions that can be measured along multiple dimensions (referred to as *parameters* by Blau). Two dimensions are consolidated when they have a positive correlation across members of a population (Blau, 1977). Consolidation creates homophily when the mechanism(s) producing homophily on one dimension lead to homophily on the second dimension *only* because of the correlation between the two dimensions. Thus, individuals associate with others who are similar on a characteristic only because that characteristic is correlated with another, more salient characteristic. For example, as consolidation increases, rates of intermarriage decrease for dimensions such as race, birth region, occupation, socio-economic status, and employment industry (Blau, Becker, & Fitzpatrick, 1984; Skvoretz, 1990). With a highly consolidated set of parameters, an ingroup tendency on one dimension leads to ingroup relations on all other parameters in the set.

Inbreeding homophily is also produced through organizations and institutions that draw similar types of individuals together. Such *substructures*, or *foci* (Feld, 1981, 1982), attract people who share similarities on many dimensions. Common foci include workplaces, voluntary associations, and neighborhoods (Feld, 1982; Fischer, 1982; Marsden, 1990; McPherson & Smith-Lovin, 1986, 1987). Relations drawn from such foci tend to be more homophilous than those not drawn from foci (Kalmijn & Flap, 2001) and more homophilous than expected by chance (McPherson & Smith-Lovin, 1987). Families are another common focus, though their effect on homophily differs. For instance, over half of the ties in the GSS discussion network data were kin-based (Marsden, 1987). While families are often similar on race, ethnicity, and religion, they are more heterogeneous on age and sex strictly due to demography.

Finally, inbreeding homophily can emerge because individuals prefer to associate with similar others (Skvoretz, 1983). Possible sources for this preference include communication facilitation (Rogers & Bhowmik, 1970), impressions of cognitive compatibility (Huston & Lvinger, 1978), and energy efficiency (Mayhew et al., 1995). Preference homophily can be conceptualized as the residual homophily that exists beyond that explained by baseline levels, consolidation, and substructural foci.

The present research focuses on homophily produced through substructures and individual preference. Baseline homophily and inbreeding homophily due to consolidation at the population level are controlled.

Multidimensionality

Several researchers have recognized the multidimensional nature of homophily in relations, i.e., that individuals are simultaneously drawn together or repelled from one another based upon a range of characteristics (Blau, 1977; Burt, 1990; Huckfeldt, 1983; Laumann, 1973; Popielarz & McPherson, 1995; Rogers & Bhowmik, 1970; Skvoretz, 1983). Early research on homophily primarily examined a single dimension (Feld, 1982; Hallinan & Williams, 1989; Robins, Elliott & Pattison, 2001) or multiple dimensions in sequence (Blau et al., 1984; Burt, 1990; McPherson & Smith-Lovin, 1987; Marsden, 1987; Skvoretz, 1990). Such research detailed the gradations of homophily, the relationship between social position and homophily, and how, independent of one another, some dimensions are more salient and exert a stronger influence on association than others (Marsden, 1987, 1988).

Recently, statistical models have emerged as a means to examine homophily on multiple dimensions simultaneously. Two examples are the stochastic actor-oriented model (Snijders, Steglich, & van de Bunt, 2010) and exponential random graph models (Robins, et al., 2007; Wasserman & Pattison, 1996). Using such

approaches, researchers have continued to find evidence for homophily on individual attributes, including race/ethnicity, sex, and age (Goodreau, Kitts, & Morris, 2009; Mouw & Entwisle, 2006; Schaefer, Light, Fabes, Hanish, & Martin, 2010). While such research considers multiple dimensions simultaneously, thereby controlling for consolidation, models typically do not include interactions between dimensions. This implicitly assumes that interactions between homophily dimensions are independent or unimportant. The current research relaxes the assumption of independent dimensions and explores how dimensions operate in conjunction to influence the likelihood of association.

The current research aims to identify which combinations of homophily occur more often than expected by chance. I present a set-theoretic approach to the patterning of homophily across several dimensions. I eschew measurement complexity in order to gain an understanding of the broader patterns of homophily across multiple dimensions. Given the limited research that has considered homophily multidimensionally, this approach serves as a useful starting point.

Configurations of Homophily

This research defines homophily as a “crisp set” (individuals are either similar or dissimilar) and relations are either in the set of “homophilous relations” or outside the set. To measure homophily for any pair of individuals, I measure their individual scores on a dimension and then classify their similarity. For a condition x , relations in the “ x homophily” set are similar on dimension x , while those outside the set are different on x . For example, relations between individuals who are the *same* age are members of the set “age homophily” and those between individuals of *different* ages are outside of the set “age homophily.”

The multidimensionality of homophily is measured by classifying relations as *configurations of homophily*. A configuration of homophily is the pattern of similarity across a

specified set of dimensions. For instance, spouses often share similar age and race but are opposite sex. Such relationships belong to the set “Age and Race Homophily.” This can be contrasted with same-sex friends, who most often fall into the set “Age, Race, and Sex Homophily.” Given a finite set of dimensions, any relation will fall into only one configuration, though some configurations will be more common than others. One would expect configurations to be more common as the number of dimensions exhibiting similarity increases.

Formally, a set of n dimensions along which homophily can be measured is defined as $X = \{x_1, x_2, \dots, x_n\}$. The *power set* of X , $P(X)$, contains all possible combinations of the elements of X , including the null set, \emptyset , in which no homophily is present. With n conditions there will be 2^n elements in $P(X)$. For example, with the conditions Age and Race, there will be four combinations in $P(X)$: “No Homophily,” “Age Homophily,” “Race Homophily,” and “Age and Race Homophily.” The 2^n elements of $P(X)$ are all of the combinations of similarity across those dimensions and are defined as *configurations of homophily*.

In the remainder of this paper, I develop two means to examine configurations of homophily. First, lattice visualization is introduced as a means to understand the nested nature of configurations and how a set of dyads is distributed across configurations. Second, Qualitative Comparative Analysis is used to uncover the interactions between homophily dimensions that make some configurations more prevalent than others.

DATA

Data come from the 1985 General Social Survey, which included a network module with questions about respondents’ discussion partners. Respondents were asked to identify others with whom they discussed important matters, to which they listed an average of 3.01

discussion partners (Marsden, 1987). The first five respondents named were included in follow-up questions about sociodemographic dimensions and their relation to the respondent. Sociodemographic information obtained from respondents and for each partner includes age, sex, race, education, and religion.

The unit of analysis for this research is pairs of individuals (dyads) consisting of respondent (ego) and discussion partner (alter). The 1,534 survey respondents provided information about 4,498 discussion partners. Drawing on only those cases with complete sociodemographic information for both ego and alter resulted in 3,999 dyads.

Configurations

Similarity between ego and alter was measured along five dimensions: age, sex, race, education, and religion. Homophily along the graduated (continuous) dimensions age and education was determined by whether or not the difference in values between ego and alter fell within a specified range. For the age dimension, pairs with differences of five years or less were coded as being similar. Education is not as straightforward because respondents only reported alters’ education in pre-specified ranges (1-6 years, 7-9 years, 10-12 years, high-school graduate, some college, A.A. degree, B.A. degree, or professional degree). To code education homophily these categories were rank ordered for both ego and alter and only those individuals who differed by less than two steps were considered similar.

Religion for both ego and alter was reported as Protestant, Catholic, Jewish, other, or none. If the religion, or non-religion, for both ego and alter were the same then that dyad was coded as having religious homophily; otherwise the dyad was coded as heterophilous. Homophily along the other nominal dimensions, sex and race, was coded in a similar manner. For each dimension, dissimilarity was coded as 0 and similarity as 1. Table 1 presents the distribution of dyads across all possible configurations of homophily (all

elements of the power set). With five sociodemographic dimensions there are 32 possible configurations. Given a large number of individuals who are heterogeneous along several dimensions, it is not surprising that all configurations of homophily are represented. The first row is the null set, absence of homophily on all five dimensions. The most common configuration is number 16, which includes similarity on four of the five dimensions (Sex, Race, Education, and Religion). This configuration contains 27% more relationships than the second most frequent configuration (number 32), which is the complete homophily configuration.

Configuration Frequency

To determine which configurations occur more often than expected by chance it is necessary to compare the observed number of relations in each configuration to a baseline level that represents random association. The theoretical baseline population in this research comprises all possible relations in the United States in 1985. Like many surveys of the general population, the GSS underrepresents males (Davis & Smith, 1996) and the marginal distributions of other demographic dimensions do not match U.S. population estimates. To create a representative baseline, the data were weighted to reflect the general population using U.S. census estimates from 1985 to weight by age, sex, and race and from 1990 to weight by education (U.S. Census Bureau, 1990, 2000). Respondents were not weighted by religion as the GSS comprises the best available national estimate of religious identification. Using the GSS as the basis for computing the baseline, rather than census estimates that don't include religion, has the advantage of preserving the consolidation of parameters. Maintaining the correlation of dimensions within individuals is vital when examining how multiple dimensions simultaneously draw people together. An expected proportion was computed for every configuration of homophily and can be seen in the Baseline column of Table 1 (computation of the baseline is presented in the Appendix).

Examination of Table 1 reveals that the distribution of dyads across configurations varies from that expected by chance. For instance, the top row represents relations that are different on all five dimensions. Examining the "All" column we see that the 0.1% of dyads observed in this configuration is well below the 3.1% expected by chance. Conversely, the 12.8% of dyads in the bottom row, homophily on all dimensions, is well above the 1.6% expected by chance. To evaluate how dimensions interact to affect the frequency of observed dyads, each configuration was coded as exceeding the baseline frequency or not. This was determined using a z -test for proportions with an alpha of .05. Those configurations whose proportion was significantly higher than the baseline proportion were coded as 1; otherwise they were coded as 0. Ten configurations occurred more often than expected by chance while twenty-two did not.

Role relations correspond to common foci that bring people together and provide the opportunity for relationships to develop (Marsden, 1990). Examining role relations independently can provide insight to the multidimensional nature of homophily induced through substructures. The GSS allowed respondents to select multiple ways in which they were connected to each alter. This research utilizes the following relations: kin, coworkers, friends, and group members. Since kin relations evolve through a different process than most other relations a distinction was made between kin and nonkin. Relations with parents, siblings, children, and other family members were collapsed into the kin relation. To help ascertain the homophily induced by substructural foci and not pre-existing kin relations, only nonkin friends, coworkers, and group members were considered. Table 1 reports the distribution of dyads across homophily configurations by type of role relation. The proportion of dyads of each relation type in each configuration was compared to the baseline, with those exceeding

Table 1. Distribution of Relations across Homophily Configurations by Role Relation

Configuration Number	Dimensions of Homophily [†]							N (3,999)	Baseline	Proportion of Relations by Type					
	Age	Sex	Race	Education	Religion	All (N=3,999)	Kin (N=1,647)			Nonkin (N=2,352)	Friend (N=1,934)	Coworker (N=603)	Group (N=530)		
1	0	0	0	0	0	0	2	.031	.001	.001	.000	.001	.000	.000	
2	0	0	0	0	0	0	7	.029	.002	.003	.001	.001	.000	.000	
3	0	0	0	0	0	0	4	.023	.001	.000	.002	.002	.003	.000	
4	0	0	0	0	0	0	13	.021	.003	.004	.003	.003	.002	.002	
5	0	0	0	0	0	0	65	.091	.016	.018	.015	.016	.022	.009	
6	0	0	0	0	0	0	287	.070	.072	.114*	.043	.039	.030	.064	
7	0	0	0	0	0	0	96	.077	.024	.024	.024	.025	.017	.009	
8	0	0	0	0	0	0	366	.065	.092*	.124*	.069	.058	.043	.066	
9	0	0	0	0	0	0	7	.030	.002	.001	.003	.003	.005	.004	
10	0	0	0	0	0	0	26	.029	.007	.009	.005	.006	.012	.000	
11	0	0	0	0	0	0	19	.022	.005	.001	.007	.006	.020	.006	
12	0	0	0	0	0	0	22	.022	.006	.009	.003	.004	.005	.004	
13	0	0	0	0	0	0	119	.090	.030	.032	.028	.031	.048	.015	
14	0	0	0	0	0	0	376	.070	.094*	.156*	.051	.049	.043	.058	
15	0	0	0	0	0	0	172	.078	.043	.047	.040	.040	.088	.019	
16	0	0	0	0	0	0	650	.065	.163*	.252*	.100*	.102*	.154*	.100*	
17	0	0	0	0	0	0	4	.007	.001	.000	.002	.001	.002	.002	
18	0	0	0	0	0	0	5	.006	.001	.000	.002	.002	.002	.002	
19	0	0	0	0	0	0	7	.006	.002	.000	.003	.003	.007	.002	
20	0	0	0	0	0	0	19	.006	.005	.001	.008	.006	.003	.006	
21	0	0	0	0	0	0	50	.019	.013	.002	.020	.020	.018	.015	
22	0	0	0	0	0	0	162	.014	.041*	.015	.059*	.048*	.033*	.077*	
23	0	0	0	0	0	0	114	.021	.029*	.004	.045*	.044*	.027	.019	
24	0	0	0	0	0	0	388	.016	.097*	.030*	.144*	.124*	.081*	.192*	
25	0	0	0	0	0	0	5	.006	.001	.000	.002	.002	.003	.000	
26	0	0	0	0	0	0	8	.006	.002	.001	.003	.004	.005	.004	
27	0	0	0	0	0	0	15	.007	.004	.001	.006	.006	.012	.004	
28	0	0	0	0	0	0	37	.006	.009*	.005	.012*	.013*	.013*	.008	
29	0	0	0	0	0	0	80	.019	.020	.006	.030*	.033*	.045*	.028	
30	0	0	0	0	0	0	154	.014	.039*	.018	.053*	.060*	.040*	.058*	
31	0	0	0	0	0	0	209	.021	.052*	.022	.074*	.085*	.075*	.070*	
32	0	0	0	0	0	0	511	.016	.128*	.103*	.145*	.163*	.144*	.157*	

[†] 0=Dissimilarity on dimension, 1=Similarity on dimension
 * Proportion exceeds the baseline proportion (z-test, $\alpha < .05$)

the baseline coded as having the outcome (occurring more often than expected by chance).

Lattice Visualization

Lattices help convey the distribution of dyads across configurations as well as the nested relationship between configurations. If X is the set of characteristics for which similarity is measured then a *partial order* can be defined on $P(X)$ using the binary relation “ \leq ” that is reflexive, antisymmetric, and transitive. Following Wasserman and Faust (1994) if c is an element of $P(X)$ then, the “ \leq ” relation can be formally stated as:

- $c_i \leq c_j$
- $c_i \leq c_j$ and $c_j \leq c_i$ if and only if $c_i = c_j$
- $c_i \leq c_j$ and $c_j \leq c_k$ implies $c_i < c_k$

Two configurations, c_i and c_j have a *lower bound* c_k when $c_k \leq c_i$ and $c_k \leq c_j$. The element c_k is the *greatest lower bound*, if $c_l \leq c_k$ for all lower

bounds c_l of c_i and c_j . Two configurations c_i and c_j have an *upper bound* c_k when $c_i \leq c_k$ and $c_j \leq c_k$. The element c_k is the *least upper bound* if $c_k \leq c_l$ for all upper bounds c_l of c_i and c_j . A *lattice* is defined as a partially ordered set of elements in which any two elements have both a least upper bound and greatest lower bound (Birkhoff, 1940). A lattice can be displayed in a diagram with nodes representing sets of elements (configurations) and lines representing the “ \leq ” relation (see Figure 1).

This lattice represents the set of five dimensions along which homophily was measured and directly corresponds to the N column in Table 1. The nodes refer to each possible configuration of homophily. Letters refer to the dimensions of similarity that comprise the configuration (for ease of presentation, the logical “and” that lies between each element of the configurations is not displayed). The uppermost node (ASREG) represents complete homophily, presence of

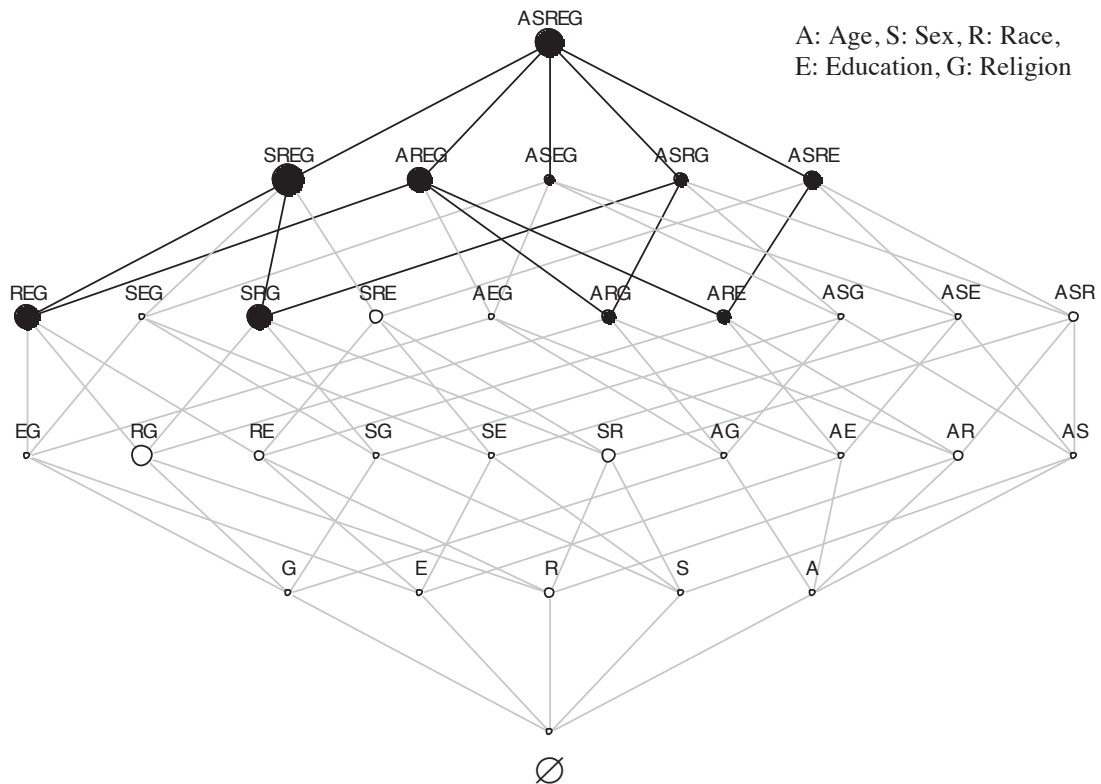


Figure 1. Distribution of Dyads Across Configurations (N=3,999)

education & religion). The lowermost node is the null set, representing complete heterophily, absence of similarity on all dimensions. Intermediate nodes represent varying degrees of homophily, increasing as one moves up the lattice. For example, the five nodes immediately below the top node represent the five configurations that differ from complete homophily on a single dimension (i.e., similarity on four conditions). The ten nodes on the level below all contain configurations with similarity on exactly three dimensions.

In this lattice, nodes are sized to indicate the frequency of a configuration within the data. The largest nodes are in the top half of the lattice, indicating that most of the dyads in the GSS sample exhibit homophily on three or more dimensions. The largest nodes are GASRE and GSRE, which combined represent over one-quarter of the observed dyads.

The placement of nodes and lines connecting them represent the nested structure of configurations in which some configurations are subsets of others. A configuration is a subset of another if all of its elements are contained within the other configuration. Two configurations are connected in the lattice if they differ on only a single element. For example, the ASREG and SREG nodes are connected because the latter is a subset of the former, differing only on the absence of age similarity. Each node contains one fewer element of similarity than the upwards node(s) to which they are connected. Nodes with multiple lines emerging from the top are subsets of multiple other configurations. Likewise, nodes with multiple lines emerging from below are supersets of multiple other configurations.

The shading in Figure 1 reflects which configurations exceed chance occurrence. Black nodes represent configurations that are more prevalent than the baseline probability and white nodes represent configurations at or below the baseline level. Note that the black nodes are concentrated at the top of the lattice, indicating that homophily on 3-5 dimensions is necessary for associations to exceed baseline rates. To better convey the set-theoretic relationship between

configurations, dark lines are used to connect configurations that exceed chance. The complete homophily configuration and the five configurations below it all occurred more often than expected by chance. Looking further down the lattice, only four additional configurations exceeded the baseline. Interestingly, none of the configurations that are subsets of ASEG (similarity on all dimensions except Race) exceed the baseline. That is, dissimilar race dyads only exceeded chance expectations when homophily was present on all four of the other dimensions. This conveys the importance of race in structuring relationships. Figures 2-5 apply the same procedure to dyads drawn from different substructural foci.

Qualitative Comparative Analysis

The patterns of homophily represented in lattices can be summarized using Qualitative Comparative Analysis (QCA). QCA is a set-theoretic approach that uses the logic of Boolean algebra to identify the combinations of causal conditions that are subsets of an outcome (Ragin, 1987, 2000). The QCA procedure applies Boolean logic to reduce the truth table to a set of statements that describe the configurations associated with the outcome. QCA identifies when either the presence or the absence of a condition is associated with an outcome. Two features in particular assist with identifying how multiple dimensions of homophily operate in conjunction. First, QCA facilitates the search for substitutable conditions, where the presence of *either* from a pair of independent variables is sufficient to produce the outcome. For example, a combinatorial model has the power to reveal that *either* race homophily or religion homophily is necessary in relations, but not both. Second, the QCA procedure is able to identify situations where the importance of one condition is contingent upon other characteristics. Similarity on a dimension may be associated with the outcome with some combinations of other dimensions, while dissimilarity may be associated with the outcome in other combinations.

In order to analyze configuration with QCA, it is necessary to specify whether the outcome occurred for each configuration, which is defined

as the configuration having membership in the outcome set (Y). In the present study, the outcome is whether or not the configuration of homophily occurred more often than expected by chance. Configurations whose frequency exceeds chance are in the outcome set; those that equal or fall below chance rates are outside of the outcome set. Only those elements of $P(X)$ that have membership in the outcome set are members of Y . The outcome is recorded for each configuration, resulting in a *truth table* that serves as the basis of analysis (i.e., Table 1).

Arriving at a QCA solution takes place through a systematic comparison of configurations that exhibit the outcome. If two configurations with the outcome differ on only one condition then that condition is considered irrelevant and can be dropped to produce a simpler expression. For example, given two configurations “Age and Race Homophily” and “Age Homophily” that both exhibit the outcome (i.e., occur more often than expected by chance), similarity on race is irrelevant to the outcome. Dyads existed more often than by chance if they were similar on age, regardless of their similarity on race. Thus, the simpler expression “Age Homophily” is an *implicant* of “Age and Race Homophily” and “Age Homophily” because both configurations are contained within the set “Age Homophily.” Comparisons between implicants can also be made and conditions that are irrelevant can be dropped. An implicant that cannot be simplified through comparison with another implicant is a *prime implicant*. The set of prime implicants for the configurations with the outcome will contain as a subset only those configurations that exhibit the outcome. Thus, the set of prime implicants for Y provides a simplified account of the combinations of conditions where the outcome exists. A QCA solution contains the combinations of conditions that are associated with the outcome. The Quine-McCluskey algorithm provides a systematic means of comparing configurations and implicants to reduce a truth table and is incorporated within the fs/QCA software used for this research (Ragin, Drass, & Davey, 2003).

The following analysis uses QCA procedures to simplify the configurations of homophily that

exceed the baseline and produce a solution that identifies the basis of inbreeding tendencies. Using fs/QCA software (Ragin et al., 2003), a crisp set analysis was performed to determine the types of configurations with more dyads than expected by chance. In the QCA models that follow, the outcome is present when the frequency of dyads in a given configuration exceeds the baseline. To simplify presentation, solutions are described as leading to the outcome or not, with solutions leading to the outcome indicating more dyads than expected by chance.

RESULTS

The results for models predicting relations of any type are presented in Table 2. The QCA solution can be interpreted as follows. The first column includes the set of solution terms from which all solutions are drawn. In each term, characteristics are connected by logical “and” (represented by *). Characteristics in uppercase letters refer to similarity on a dimension while lowercase letters refer to dissimilarity. The columns to the right indicate which terms are contained in the solution for each type of relationship (designated with a “•”). The terms for each solution are connected by logical “or.” Thus, the first solution (column 2) indicates that more dyads than expected by chance existed when ego and alter had similar:

sex and race and religion, or
age and race and religion, or
race and education and religion, or
age and race and education, or
age and sex and education and religion

The strong effect of race homophily is evident in the solution. All terms in the solution term, include similarity on race except for the final which includes similarity on every dimension except for race. Religion is also quite strong in that four of the terms include similarity on religion. However, the number of dyads with just similarity on race and religion did not exceed chance. It was necessary to combine similarity on race and religion with at least one other dimension in order to produce inbreeding homophily.

This solution can be better understood by examining the lattice in Figure 1. Because all

configurations whose frequency exceeded the baseline (i.e., the black nodes) are no more than two levels down from the top node, all configurations with the outcome contain similarity on at least three dimensions. The lowermost terms that contain the outcome correspond to the components of the QCA solution. For example, the solution term (AGE*RACE*EDUC) indicates that configurations that were similar on these three dimensions contained more dyads than expected by chance. This solution term corresponds to the node labeled “ARE” and all supersets of this node. That is, solution terms may *imply*, or explain, the outcome in multiple configurations. Configurations that are supersets of the solution term can be identified by tracing

the lines upward. When solution terms only contain presence on conditions then all nodes configuration is a subset of two configurations with greater similarity (“ASRE” and “GARE”), which are subsets of the complete homophily configuration. Thus, the solution term “AGE*RACE*EDUC” implies the four configurations from “ARE” upward.

Now consider the solution term (AGE*SEX*EDUC*RELG), which contains four dimensions of similarity. This term corresponds to the node labeled “ASEG.” This solution term cannot be simplified because all nodes below represent configurations whose frequency did not exceed the baseline threshold. The four terms

Table 2. QCA Solutions by Relationship Type[†]

Solution Terms	All Dyads	Kin	Nonkin	Nonkin Only		
				Friend	Coworker	Group Member
SEX * RACE * RELIG	●					
AGE * RACE * RELIG		●				
AGE * RACE * RELIG	●		●	●	●	●
RACE * EDUC * RELIG	●	●				
SEX * RACE * EDUC * RELIG			●	●	●	●
AGE * SEX * RACE			●	●	●	
AGE * RACE * EDUC	●		●	●		
AGE * SEX * EDUC * RELIG	●		●	●	●	
AGE * SEX * RACE * EDUC						●

[†] Uppercase refers to similarity, lowercase to dissimilarity, * to the logical operation "and"

below represent configurations that are dissimilar on only two dimensions, race and one other dimension, yet none of those configurations exceeded the baseline frequency. This result emphasizes that dissimilarity on race inhibits relationships unless individuals are similar along the four other dimensions.

The configurations and outcomes for kin and nonkin are presented in Figures 2 and 3 respectively. The corresponding results of the QCA analysis performed on each subset are presented in columns 3 and 4 of Table 2.

The QCA solution for kin consists of two components, both of which contain similarity on race and religion. Two solution terms are required because two of the configurations that contain similarity on race and religion did not exhibit the outcome (“ARG” and “ASRG”). In order to exclude these configurations from the QCA solution it was necessary to include dissimilarity on age in one of the solution terms, which can be seen in the first solution term. The first solution component contains intergenerational relationships, likely between parents and children, aunts/uncles and nieces/nephews, or grandparents/grandchildren.

The types of homophily in nonkin relations are quite different from kin relations. This contrast is most evident when comparing Figure 2 with Figure 3. Of the nine nonkin configurations with the outcome, only three were also associated with more kin than expected by chance. This difference is also evident in the QCA solution. Three of the nonkin QCA solution components contain age and race similarity combined with similarity on one other dimension, indicating that similarity on age and race was important, but required similarity in one more aspect for inbreeding homophily to occur. This is shown in Figure 3 by the three black nodes on the right side of the lattice (“ARG,” “ARE,” and “ASR,”). Each of these nodes requires its own solution term because they do not share a common lower bound that exhibits the outcome. Each of these

configurations is a superset of the “AR” configuration, which includes dyads with age and race similarity only. But, the frequency of this configuration does not exceed the baseline, demonstrating that similarity on age and race is not sufficient to produce more relations than expected by chance. Similarity on one more dimension must exist.

The other two nonkin solution components (“ASEG” and “SREG”) include similarity on four of the five dimensions. Both solution components contain similarity on sex, education, and religion with additional similarity on either age or race required. Combined with the first three solution components, this highlights the importance of race and age in nonkin relations. Dyads that differed on race or age only exceeded chance rates when homophily existed on all four other dimensions.

The same analysis was conducted separately for nonkin friends, coworkers, and group members. The first relation analyzed was friend, which can include friends as well as romantically involved partners who are not married. The results for the QCA solution are presented in the fifth column of Table 2 and are identical to the solution obtained for all nonkin relations. This common finding is not surprising given that friends are the largest category of nonkin relations (82% of nonkin alters were identified as friends). As noted above, this solution highlights the importance of age and race homophily. Three of the solution components combine homophily on age and race with one other dimension. The other two solution terms contain dissimilarity on age and race respectively, but similarity on the other four dimensions. Note that two of the three terms include dissimilarity on education combined with similarity on age and race. Since the close friends identified in the GSS often have long-standing relationships (Burt, 1990) it is possible that these were childhood friendships that persisted despite individuals completing their education at different points.

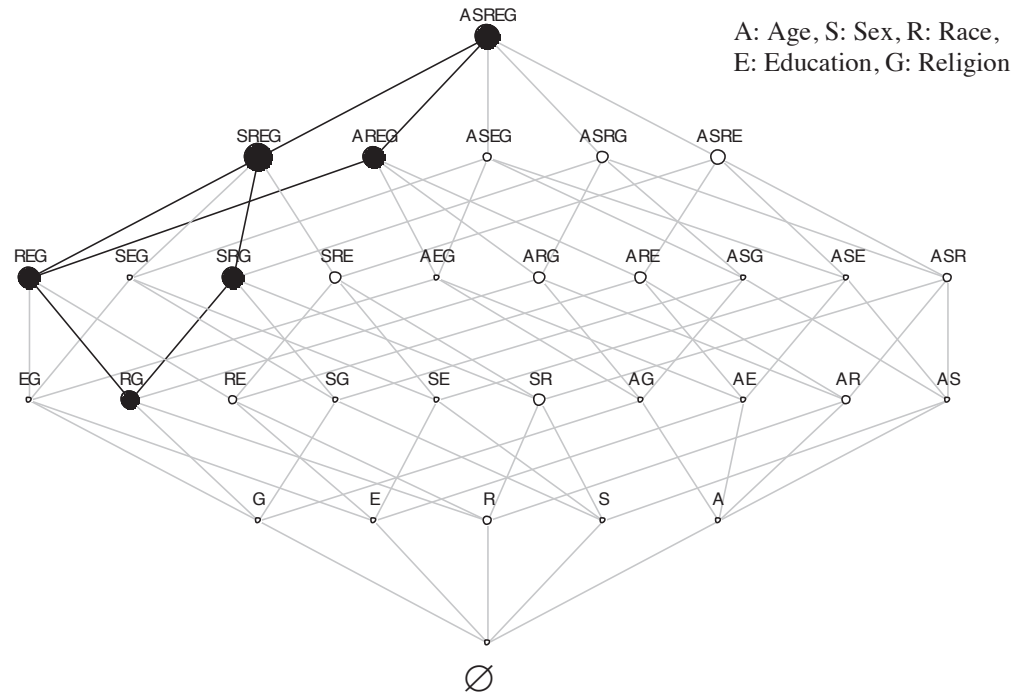


Figure 2. Distribution of Outcome Across Configurations, Kin Relations Only (N=1,647)

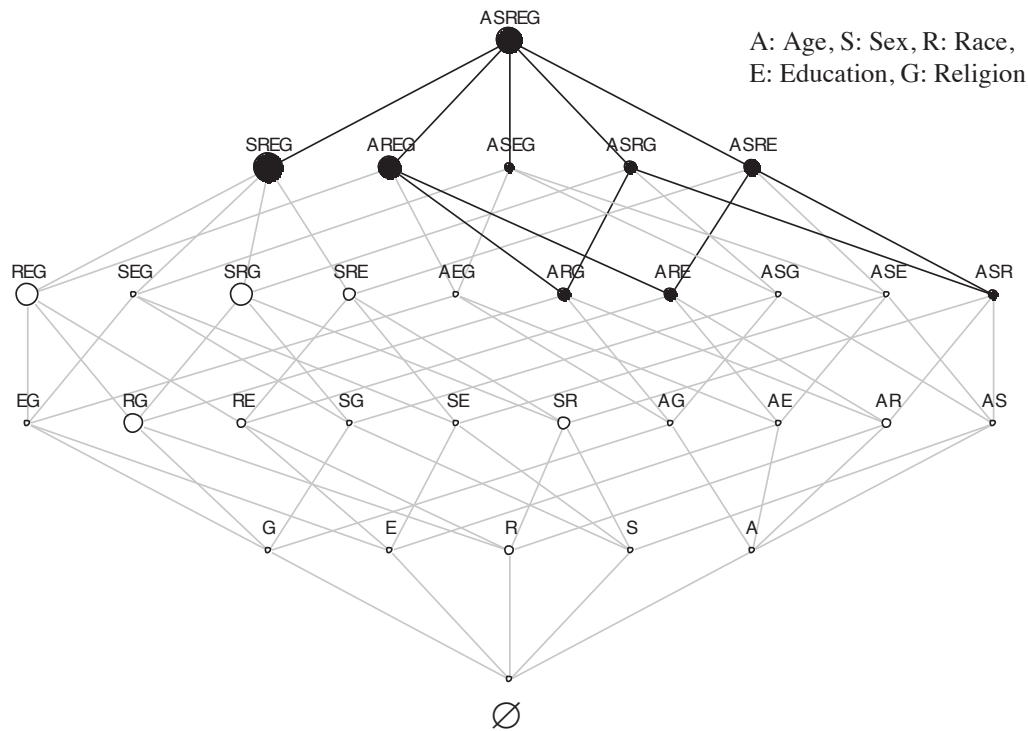


Figure 3. Distribution of Outcome Across Configurations, Nonkin Relations Only (N=2,352)
Same Distribution as Friend Relations (N=1,934)

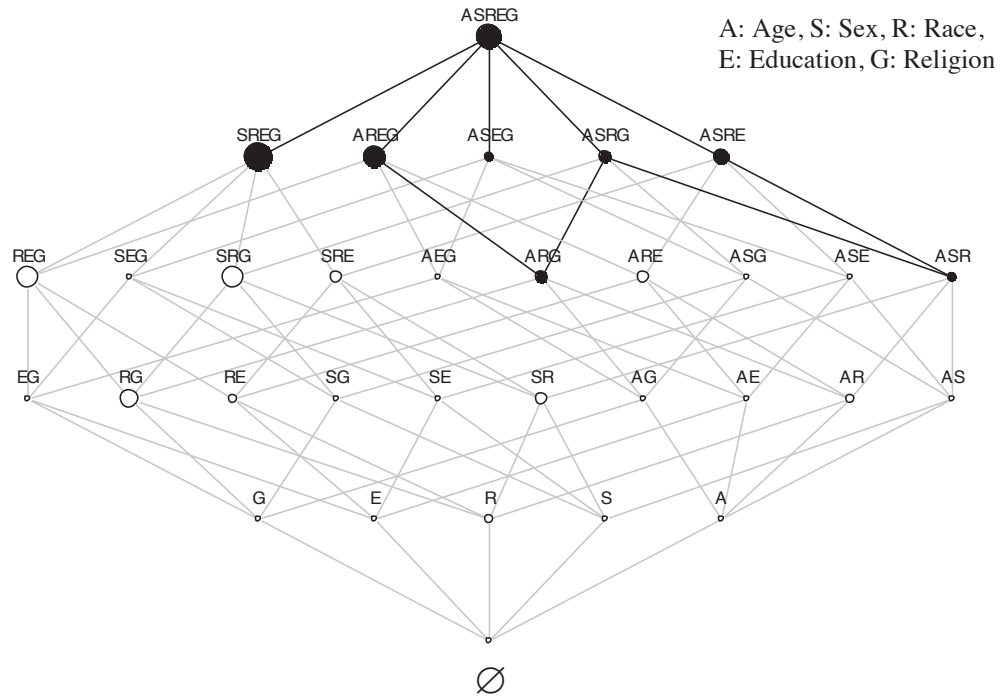


Figure 4. Distribution of Outcome across Configurations, Coworker Relations (N=603)

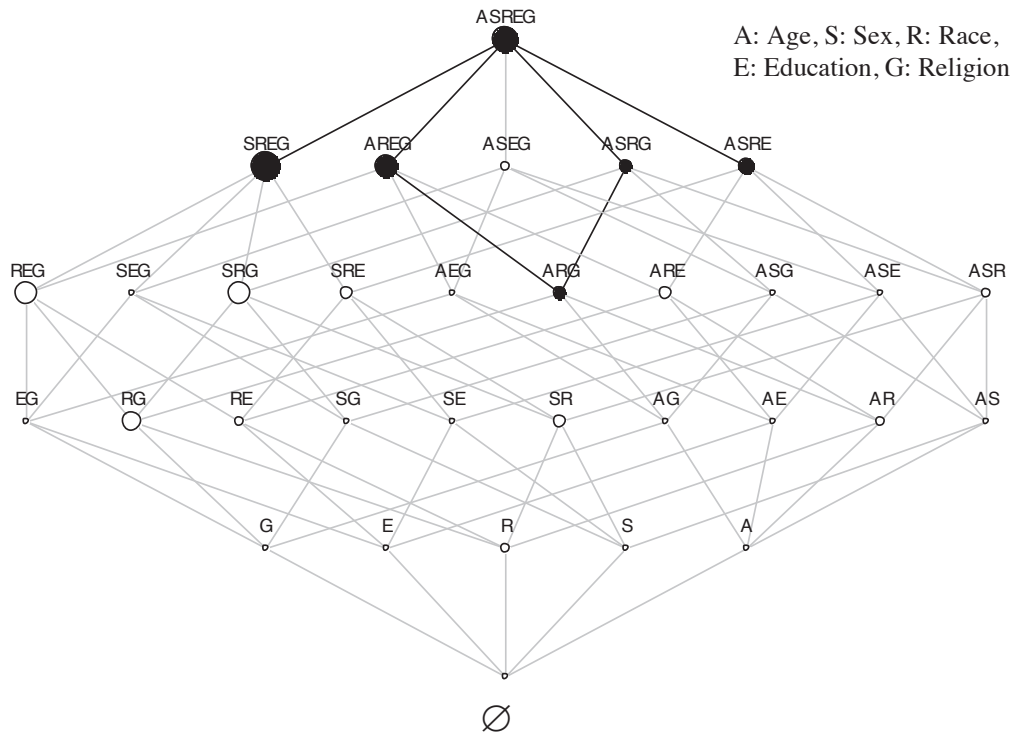


Figure 5. Distribution of Outcome across Configurations, Group Member Relations (N=530)

Homophily among relations with coworkers was examined next. The distribution of dyads across configurations was similar to that of friends with one exception. Unlike friendship relations, relations with coworkers differing only on religion and sex did not exceed the baseline level. This can be seen easily in the lattice in Figure 4 where the “ARE” node is now white, indicating the outcome is absent. The QCA solution component that explained this configuration in the friendship relation (AGE*RACE*EDUC) was dropped because it was not necessary to explain this configuration or any configurations above that node. Homophily on age and race was again key in relations between coworkers. Dyads that differed on one of these two dimensions only exceeded baseline levels when all other dimensions were similar. Coworkers also tended to be highly homophilous, with dissimilarity on education the most noticeable departure. The effect of education may be due to greater education diversity in the workplace. Because individuals tend to have friends at work, higher rates of education heterogeneity in the workplace should decrease education homophily in relations (Blau 1977).

The pattern of homophily among group members is displayed in Figure 5. Consistent with prior research on voluntary associations (McPherson and Smith-Lovin 1987), relations between group members tended to be homophilous. The frequency of these dyads exceeded the baseline in fewer configurations than with friendships or coworkers – six configurations rather than eight. This can be seen in Figure 5 where the “ASR” and “ASEG” configurations no longer exceed baseline levels. This is also reflected in the simpler, three-component QCA solution shown in Table 2.

Among group members, age, sex and race were the primary inbreeding dimensions. Similarity on race is present in all configurations that exceed the baseline, while age and sex similarity exists in all configurations but one, in each case where similarity on all other dimensions exists. This is to be expected as voluntary associations offer more choice in membership than the workplace. Only one configuration differed on more than a

single dimension (“ARG”), and one configuration that is similar on four dimensions (“ASEG”) still did not exceed the baseline level due to the significance of race.

DISCUSSION

Homophily is one of the primary forces operating in social networks; however, all too often it has been treated as a unidimensional construct. This research investigated how dimensions of homophily might operate in conjunction to affect relationship frequency. I adopted a configurational approach in order to gain initial insight to the association between five dimensions of homophily. First, lattices were introduced as a means of visualizing the set-theoretic nature of homophily. Lattices are a valuable means to represent the associations between cases with configurational data and facilitate the comparison of cases along different outcomes (in this case, relation types). Second, QCA was applied to configurations to understand the combinations of homophily necessary for dyads to exceed baseline levels. Together, these tools offered new insights to the nature of homophily.

The analysis revealed several broad patterns. To begin, a considerable degree of homophily was necessary for dyad frequency to exceed baseline levels. While 87% of relations differed on at least one of the five dimensions, in only one case was this due to anything beyond baseline expectations (age heterophily for family members). In general, a configuration must have included similarity on at least three of the five dimensions to exceed chance expectations. The sole exception was kin relations, which exceeded baseline levels with similarity on race and religion only because age dissimilarity was more common. At the same time, complete homophily was not necessary. Several configurations displaying *dissimilarity* on one or two dimensions exceeded baseline levels. This suggests that similarity on some combinations of dimensions may overcome dissimilarity on other dimensions. That is, homophily on the right set of dimensions may be enough to draw people together beyond chance rates.

These results offer support to prior findings on the relative importance of dimensions. Relations that were homophilous on only four dimensions typically exceeded baseline levels, however race was a consistent exception. For kin and group member relations notably, configurations did not exceed baseline levels unless they contained similarity on race. It was necessary to combine similarity on race with other dimensions for dyad frequency to exceed chance levels, pointing to the influential role of race in structuring relations. Friends and coworkers were the exception to this tendency as race dissimilar relations exceeded baseline rates, but only when similarity on all other dimensions was present.

Interesting patterns were found across the different types of role relations. All kin relations that exceeded the baseline level included similarity on both race and religion. However, greater age dissimilarity was seen in kin versus nonkin relations, with half of the kin configurations exceeding baseline levels exhibiting age dissimilarity. This is a function of the substructure of families, where most members are similar on race and religion, though with greater variability in age, sex, and education. By contrast, in nonkin relations, religion was less important, and race was just as important though in a different way. Most nonkin relations that exceeded the baseline level included similarity on both age and race, combined with homophily on a third dimension. That third dimension varied by role; with group members, religion accompanied age and race as homophilous dimensions; with friends and coworkers greater variety existed. Note that for no type of nonkin relation did a configuration with dissimilarity on race and another dimension or age and another dimension exceed the baseline level. Relations with race or age dissimilarity combined with another dimension were too different to exceed the baseline level.

Across all relation types, ten of the thirty-two configurations exceeded baseline levels. Examining results by substructural foci reveals their effect on the development of intergroup relations. The greatest diversity in homophily configurations was seen in coworker relations,

where eight different configurations exceeded baseline levels. In contrast, only six configurations of homophily exceeding baseline levels for group relations. This is an indicator of higher levels of homogeneity in voluntary association relations. In addition, though the pattern of homophily differs for kin relations, they were just as homogeneous as groups, with only six configurations exceeding the baseline level. These results are consistent with Marsden's study of network diversity (1990).

One limitation of this research was the treatment of homophily and configurations as present or absent. The current approach introduced a technique to visualize the pattern of configurations using lattices, though at the expense of measurement precision. One could extend this research by using fuzzy set measurement techniques (where homophily could be measured continuously from 0 to 1). Such an approach could capture gradations in homophily, whereby two individuals are not identical, but not completely dissimilar. Subsequently, configuration frequency could also be measured using fuzzy set techniques, which could incorporate measures of relationship closeness including strength, duration, time spent together, and multiplexity. Measuring the outcome as a fuzzy set may provide insight to those instances where configuration frequency exceeds baseline levels even though dissimilarity on multiple dimensions exists. Perhaps those relationships are not as close as those with greater levels of homophily. Either approach would build on the insights gleaned from this research.

In addition, this research treated relations as the unit of analysis without regard for the social positions of the individuals. Prior research has demonstrated how network characteristics vary according to social position (Fischer, 1982; Ibarra, 1995; Marsden, 1987; Moore, 1990). Further research on the multidimensionality of homophily should consider variation across social positions. If network heterogeneity is associated with social mobility as suggested by Blau (1977) and social capital as Lin suggests (2001) then examining the coincidence of homophily across multiple dimensions by social position can provide clearer

insight to the structures that preserve stratification.

This research used ego network data to find that dimensions of homophily interact in patterned ways to structure network relations. Tools allowing the simultaneous measurement of homophily on multiple dimensions are becoming

increasingly available for *complete* network data (Snijders et al., 2010). Thus, it is now common for statistical models to test for homophily on several dimensions when modeling network structures. Such models could easily be used to extend this research by testing which dimensions operate conjointly in relations drawn from different substructures.

REFERENCES

- Birkhoff, G. 1940. *Lattice Theory*. Providence: American Mathematical Society
- Bienstock, E., P. Bonacich, and M. Oliver. 1990. "The effect of network density and homogeneity on attitude polarization." *Social Networks* 12:153-72.
- Blau, P. 1977. *Inequality and Heterogeneity: A Primitive Theory of Social Structure*. New York: Free Press.
- Blau, P.M. and J.E. Schwartz. 1984. *Crosscutting Social Circles*. Orlando, FL: Academic Press.
- Blau, P.M., Becker, C. and K.M. Fitzpatrick. 1984. "Intersecting Social Affiliations and Intermarriage." *Social Forces* 62:585-606.
- Blum, T.C. 1985. "Structural Constraints on Interpersonal Relations: A Test of Blau's Macrosociological Theory." *American Journal of Sociology* 91:511-21.
- Burt, R. 1990. "Kinds of Relations in American Discussion Networks." Pp 411-51 in *Structures of Power and Constraint: Papers in Honor of Peter Blau*, Ed. C Calhoun, M Meyer, and R Scott. New York: Cambridge.
- Buston, P. M. & Emlen, S. T. 2003 Cognitive processes underlying human mate choice: The relationship between self-perception and mate preference in Western society. *Proceedings of the National Academy of Science*, 100, 8805-10.
- Davis, J.A. and T.W. Smith. 1996. *General Social Surveys, 1972-1996: Cumulative Codebook*. Chicago: National Opinion Research Center.
- Feld, S. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology*. 86:1015-35.
- Feld, S.L. 1982. "Social Structural Determinants of Similarity among Associates." *American Sociological Review* 47:797-801.
- Fischer, C. 1982. *To Dwell Among Friends*. New York: Free Press.
- Freeman, L.C. and D.R. White. 1993. "Using Galois Lattices to Represent Network Data." *Sociological Methodology* 23:127-46.
- Goodreau, S. M., Kitts, J. A., & Morris, M. (2009). Birds of a feather, or friend of a friend?: Using exponential random graph models to investigate adolescent social networks. *Demography*, 46, 103-125.
- Hallinan, M and R Williams. 1989. "Interracial Friendship choice in Secondary Schools." *American Sociological Review*. 54: 67-78.
- Huckfeldt, R.R. 1983. "Social Contexts, Social Networks, and Urban Neighborhoods: Environmental Constraints on Friendship Choice." *American Journal of Sociology* 89:651-69.
- Huston, T.L. and G. Levinger. 1978. "Interpersonal Attraction and Relationships." *Annual Review of Psychology*. 29:115-56.

- Ibarra, H. 1992. "Homophily and Differential Returns: Sex Differences in Network Structure and Access in an Advertising Firm." *Administrative Science Quarterly* 37:422-47.
- Ibarra, H. 1995. "Race, Opportunity, and Diversity of Social Circles in Managerial Networks." *Academy of Management Review*. 38: 673-703.
- Kalmijn, M. 1998. "Intermarriage and Homogamy: Causes, patterns and Trends." *Annual Review of Sociology*. 24: 395-421.
- Kalmijn, M. and H. Flap. 2001. "Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices." *Social Forces* 79:1289-1312.
- Laumann, E.O. 1963. *Bonds of Pluralism*. New York: Wiley.
- Lin, N. 2001. *Social Capital: A Theory of Social Structure and Action*. Cambridge: Cambridge University Press.
- Louch H. 2000. "Personal Network Integration: Transitivity and Homophily in Strong-Tie Relations." *Social Networks* 22: 45-64.
- Marsden, P.V. 1981. "Models and Methods for Characterizing the Structural Parameters of Groups." *Social Networks* 3:1-27.
- Marsden, P. 1987. "Core Discussion Networks of Americans." *American Sociological Review*. 52: 122-131.
- Marsden, P. 1988. "Homogeneity in Confiding Relations." *Social Networks*. 10: 57-76.
- Marsden, P. 1990. "Network Diversity, Substructures for Opportunity and Contact." Pp 397-410 in *Structures of Power and Constraint: Papers in Honor of Peter Blau*, Ed. C Calhoun, M Meyer, and R Scott. New York: Cambridge.
- Mayhew, B.H., McPherson, M., Rotolo, T. and L. Smith-Lovin. 1995. "Sex and Ethnic Heterogeneity in Face-to-Face Groups in Public Places: An Ecological Perspective on Social Interaction." *Social Forces* 74:15-52.
- McPherson, J.M. and L. Smith-Lovin. 1986. "Sex Segregation in Voluntary Associations." *American Sociological Review* 52:370-79.
- McPherson, M and L Smith-Lovin. 1987. "Homophily in Voluntary Organizations: Status Distance and the Composition of Face-to-Face Groups." *American Sociological Review*. 52:370-9
- McPherson, M, L Smith-Lovin, and J Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415-44.
- Mouw, T., & Entwisle, B. (2006). Residential segregation and interracial friendship in schools. *American Journal of Sociology*, 112, 394-441.
- Popielarz, P.A. and J.M. McPherson. 1995. "On the Edge or In Between: Niche Position, Niche Overlap, and the Duration of Voluntary Association Memberships." *American Journal of Sociology* 101:698-720.
- Ragin, C. 1987. *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley: University of California Press.
- Ragin, C. 2000. *Fuzzy-Set Social Science*. Chicago: University of Chicago Press.
- Ragin, Charles C., Kriss A. Drass and Sean Davey. 2003. *Fuzzy-Set/Qualitative Comparative Analysis 1.1*. Tucson, Arizona: Department of Sociology, University of Arizona.
- Robins, G., Elliott P. and P. Pattison. 2001. "Network Models for Social Selection Processes." *Social Networks* 23: 1-30.
- Robins, Garry, Tom A.B. Snijders, Peng Wang, Mark Handcock, & Phillipa E Pattison. 2007. Recent developments in exponential random graph (p*) models for social networks. *Social Networks* 29, 192-215.
- Rogers, Everett M., and Phillip K. Bhomik. 1970. "Homophily-heterophily: Relational concepts for communication research." *Public Opinion Quarterly* 34: 523-538.

- Schaefer, D. R., Light, J. M., Fabes, R. A., Hanish, L. D., & Martin, C. M. (2010). Fundamental principles of network formation among preschool children. *Social Networks*, 32, 61-71.
- Skvoretz, J. 1983. "Salience, Heterogeneity, and Consolidation of Parameters: Civilizing Blau's Primitive Theory." *American Sociological Review* 48:360-75.
- Skvoretz, J. 1990. "Biased Net Theory: Approximations, Simulations and Observations." *Social Networks* 12:217-38.
- Steglich, Christian, Tom A. B. Snijders, & Michael Pearson. (2010). Dynamic networks and behavior: Separating selection from influence. *Sociological Methodology*, 20, 1-65.
- Travers, J. and S. Milgram. 1969. "An Experimental Study of the Small World Problem." *Sociometry* 32:425-443.
- U.S. Census Bureau. 1990. *1990 Census of Population: Education in the United States*. Washington D.C.: U.S. Government Printing Office.
- U.S. Census Bureau. 2000. "U.S. Population Estimates by Age, Sex, Race, and Hispanic Origin: 1980 to 1999." http://www.census.gov/popest/archives/1980s/80s_nat_detail.html
- Wasserman, S. and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.
- Wasserman, Stanley and Phillipa E. Pattison. 1996. Logit models and logistic regression for social networks I. An introduction to Markov graphs and p*. *Psychometrika* 61, 401-425.
- Watts, D.J. 1999. "Networks, Dynamics and the Small-World Phenomenon." *American Journal of Sociology* 105:493-527.

Appendix: Baseline Computation

To compute baseline homophily, the population distribution must be measured across a set of parameters. One can then compute how often random pairs of individuals drawn from the population would be homophilous. For example, if a population is 70% male and 30% female, then the expected proportion of male-male relations would be .49 (.7 x .7), female-female relations would be .09 (.3 x .3), male-female and female-male relations would be .21 each (.7 x .3). Thus, 58% of relations are expected to be homophilous and 42% of relations heterophilous. For a single nominal dimension with i categories the expected proportion of homophilous relations M is:

$$[1] M_n = \sum_i p_i^2$$

This formula can also be used to compute the expected proportion of homophilous ties for graduated, or interval level, parameters if homophily is conceptualized strictly as absolute similarity. Often however, homophily is defined using a range, such that individuals whose values on a dimension are within a specified distance are considered similar. For example, age homophily can be defined as two individuals with the exact same age or as two individuals with ages that fall within five years of one another. The following formula can be used to measure homophily for a single graduated parameter using the broader definition:

$$[2] M_g = \sum_i p_i p_{k(i)}$$

where $p_{k(i)}$ is defined as the proportion of cases falling within a specified range around i that is defined as homophilous. Using the example of age and a width of \pm five years, if i is 37 then k is ages 32-42.

For multiple parameters, possible consolidation requires that the computation of baseline homophily incorporate the joint distribution of individuals across dimensions. The marginal distributions across dimensions become irrelevant and the joint distribution is utilized in calculating the baseline. For two nominal dimensions with i and j categories respectively, baseline homophily is computed as:

$$[3] M_{n2} = \sum_i \sum_j p_{ij}^2$$

For two graduated parameters baseline computation must incorporate the range around values of each dimension, which are often distinct. The equation then becomes:

$$[4] M_{g2} = \sum_i \sum_j (p_{ij}) p_{k(i)} p_{l(j)}$$

Here, l refers to the range around the value j of the second graduated parameter that is considered homophilous with j .

Finally, equations 1 and 2 can be extended to compute baseline homophily for one nominal and one graduated dimension together as follows:

$$[5] M_{ng} = \sum_i \sum_j p_{ij} p_{k(i)} p_j$$

Through a similar extension process, baseline homophily can be computed for multiple dimensions simultaneously. This research utilized the following equation for three nominal dimensions (sex, race, religion) and two graduated dimensions (age, education).

$$[6] M_{n3g2} = \sum_s \sum_r \sum_g \sum_a \sum_e p_{srgae}^2 (p_{ae}) p_{i(a)} p_{j(e)}$$

The formulas presented above have a slight bias in small populations due to the fact that one cannot form a relationship with oneself. This bias approaches zero as population size increases and is negligible when considering a population the size of the U.S.

Parents' Networks: Egocentric Networks and Unique and Shared Sources of Social Support

Jeffrey A. Hall

*Department of Communication Studies
University of Kansas*

Abstract

Egocentric social network instruments typically require independently sampled respondents to identify up to five social network alters. When collecting egocentric data from dyads (e.g., mothers and fathers), shared and unique network alters can be identified. The present manuscript describes a new way of using egocentric data collected from related pairs using Multilevel Modeling (MLM). As a case study, the egocentric social support networks of twenty pairs of parents of children with cancer ($N = 40$) will be analyzed to illustrate how this technique can be used to model the characteristics of each network alter and to answer research questions regarding sex differences in received social support networks.

Author

Jeffrey Hall is an Assistant Professor in the Department of Communication Studies at the University of Kansas. Dr. Hall's research focuses on the influence of gender on personal relationships in several contexts. His current work focuses on sex differences in friendship expectations, homophobia and masculinity, courtship and flirting, and romantic couples' use of humor.

Acknowledgements

This manuscript includes a portion of the author's dissertation data, conducted under the direction of Dr. Michael J. Cody at the Annenberg School for Communication at the University of Southern California. Appreciation is extended to Dr. Shelia Murphy and Dr. Michael Messner for service and guidance on the dissertation committee.

Please address correspondence to Dr Jeffrey A. Hall, Department of Communication Studies, University of Kansas, Bailey Hall, 1440 Jayhawk Boulevard, Lawrence, KS., 66045-7574, Fax: 785-864-5203, e-mail: hallj@ku.edu

INTRODUCTION

Social network analysis (SNA) can document, illustrate, and identify how individuals' networks function and how members of a network interact with one another (Wasserman & Faust, 1994). Egocentric networks are formed when independently sampled individuals are asked to describe their immediate social network (Marsden, 1990). However, these data can pose difficulties for social network researchers because respondents are typically unassociated, so complete social networks cannot be constructed. This manuscript describes a new way of using egocentric data collected from dyads using Multilevel Modeling (MLM). Challenges of using egocentric data gathered from relational dyads, the utility of differentiating shared and unique alters, and the use of MLM to analyze data using the one-with-many design are explored (Kenny, Kashy, & Cook, 2006). As a case study, the social support networks of twenty pairs of parents of children with cancer ($N = 40$) will be analyzed to illustrate how this technique can be used to test theoretically-informed research questions.

Egocentric Data

Egocentric methods have a long history in SNA (Marsden, 1987, 1990). For over 25 years, the General Social Survey has used egocentric network methods to ask respondents with whom they have discussed important matters (Marsden, 1987). These data describe the number, the heterogeneity (i.e., age, education, sex), and the composition of network alters (i.e., kin vs. non-kin) (Burt, 1984; Marsden, 1987), and usefully describe social trends in the nature and composition of Americans' social networks (Marsden, 1987). Egocentric network methods have also been employed widely in received social support research (Wellman & Hiscott, 1985).

The difference between received and perceived social support is both a theoretical and methodological question. Perceived social support is defined as the perceived availability

of support if it were needed (Cutrona, 1996). Received social support is defined as tangible, measurable, and enumerable sources of various types of social support (i.e., instrumental, material, informational) (Cutrona, 1996). Received social support is often measured using egocentric network methods (Marsden, 1990; Wills & Shinar, 2000). Many received social support instruments, such as the UCLA Social Support Interview (Wills & Shinar, 2000), ask respondents to identify up to five sources of social support. In comparison to perceived social support, SNA offers unique advantages in identifying the source of support, the type of support, and the quality of support from each source (Brisette, Cohen, & Seeman, 2000; Wills & Shinar, 2000), allowing researchers to identify what types and sources of social support are most effective (Wellman & Hiscott, 1985). However, researchers often reduce the amount of variance available when using egocentric networks by combining all sources of social support into a single measure of received support (see Bissette et al., 2000). MLM offers researchers the opportunity to analyze egocentric data while preserving the unique variance of each network alter.

One-With-Many Design

Analysis of egocentric SNA data can be greatly improved by utilizing MLM. Although combining data from alters can produce useful information regarding network composition, including heterogeneity and size (see Burt, 1984; Marsden, 1987), it is not always appropriate to calculate a mean score for all alters in a social support network. A respondent with moderate or weak support across all members will have the same mean support as a respondent with a few very good quality and a few very poor quality sources of support. Every member of a support network is not equally valuable or equally supportive (Thoits, 1995), and for researchers, knowing the particular characteristics that make a source valuable is desirable. For example, to test hypotheses regarding how the quality or type of support is related to alters'

characteristics, the relative value of each alter must be tested individually rather than summed. Kenny, Kashy, and Cook (2006) identify the one-perceiver many-targets design as the most common one-with-many design. This research design asks respondents to evaluate other members of a social network, often using egocentric methods. This poses unique analysis challenges because the data are non-independent in that they share a common fate -- the relationship with the ego (Kenny et al., 2006). This design is best served by MLM in that the shared variance of the ego is modeled by treating each ego as a Level 2 predictor and network members as Level 1 observations (Kenny et al., 2006). This controls for non-independence of egocentric networks. Additionally, the unique qualities of each alter can be tested in relation to outcomes associated with the relationship with the ego. For example, this method allows for the qualities of alters (e.g., kin v. non-kin, demographic characteristics) to be estimated in relation to the outcomes of each tie (e.g., overall quality of support).

Dyadic Egocentric Data

The use of dyadic data-- that is data collected from a non-independent pair of individuals --has grown substantially in social science research (Kenny et al., 2006). In fact, some research traditions are fundamentally concerned with relationships between individuals. For example, family communication emphasizes a systems theory approach, which asserts that all family members influence one another (Segrin & Flora, 2005; Street, 2003). The theoretical emphasis and growing interest in exploring how individuals within a couple influence one another has only increased interest in collecting dyadic data. Although research has begun to illustrate the highly interconnected nature of social support for couples (e.g., Widmer, 2006), there are still many questions about how and for whom support is provided (Cutrona, 1996). Identifying characteristics about the source of support helps to overcome a gap in research on received support and can help researchers to

better understand how support can be marshaled and utilized (Thoits, 1995). Although dyadic data poses particular analytic challenges, when collecting dyadic data using egocentric network methods, there are also valuable opportunities.

When a SNA is performed on a bounded group and the identities of all network members are known, identifying shared versus unique alters is easy and is an important part of describing network characteristics (Scott, 1991). However, when collecting egocentric data from a dyad, the identity of unique versus shared sources of support can be particularly valuable both theoretically and empirically. When respondents complete an egocentric instrument, they are often asked to both provide identifying information about each network alter (e.g., name, initials) as well as information about each alter, such as sex, their relationship to the ego (e.g., step-mother), and sometimes other characteristics as well (Burt, 1984). When this information is gathered from both individuals in a dyad, their responses can be linked, and alters can be identified as shared or unique. The utility of exploring egocentric network characteristics and shared and unique sources of social support is discussed in the following case study.

Case Study: Parents' Social Support Networks

Clinical research on parents of children with cancer has revealed that social support is an important predictor of parents' health. However, Hoekstra-Weebers and colleagues consistently demonstrate that there are different social support predictors of long-term health for fathers and mothers (Hoekstra-Weebers, Jaspers, Kamps, & Klip, 1998; Hoekstra-Weebers, Jaspers, Kamps, & Klip, 2001). The differences between fathers and mothers are partly explainable by the differences in social support they receive. Fathers often lack a means to obtain the desired amount of social support, both in quality and in quantity (Hoekstra-Weebers, Jaspers, Kamps, & Klip, 1999; Hoekstra-Weebers, Jaspers, & Kamps, 2000; Sloper, 2000). There are several explanations regarding

men's lack of ability to obtain the desired, high quality support.

Differences in support network composition offer one explanation for the lack of quality and quantity of men's received support. In comparison to men, women have more sources of social support from friends and kin (Stevens & Westerhof, 2006). In the context of pediatric cancer, even when fathers have broad social support networks, they are often dissatisfied with the quality of received support (Hoekstra-Weebers et al., 2000; Hoekstra-Weebers et al., 2001). Men have fewer sources of social support and the sources they do have are not of sufficient quality. Additionally, the sex composition of parents' support networks may explain these differences. Burleson and Kunkel (2006) suggest that women are more likely than men to provide quality emotional support, and women are often more active support network members (Widmer, 2006). Clinical research has suggested that if a man has a male-dominated social support system, he may be able to draw only limited support (Hoekstra-Weebers et al., 2001). A second explanation for men's lack of quality support suggests that family members provide different types of support to mothers and fathers. During a child's treatment for cancer, mothers often serve as the primary caretaker, and as such is the focal person for family-initiated support (Sloper, 2000). Other research in non-clinical environments demonstrates that mothers often receive more support from family members than fathers (Stevens & Westerhof, 2006). In light of these research findings, there are three possible explanations for fathers' lack of support. It is possible that shared sources of support are adequately supporting mothers but fail to support fathers. On the other hand, family members may provide high quality support to both parents, but mothers have additional sources of support that fathers do not share. Additionally, it is possible that fathers and mothers are receiving different types of support that differ in relation to overall support quality. All three research questions will be answered by differentiating shared v. unique network alters for parents and by using MLM.

METHODS

Recruitment Procedure

In coordination with Childrens Hospital in a major metropolitan area in the Western US, surveys were administered in both Spanish ($N = 11$) and English ($N = 29$) during an annual survivorship festival. Participants were parents of a child who had undergone or was currently undergoing cancer treatment and were either married or living together at the time of their child's treatment. Parents who met the criteria were consented and completed the survey instrument separately. Twenty father-mother pairs participated ($N = 40$).

Participants

Fathers were 41 years old on average ($SD = 7.86$, range 31-56), and 95% were employed full time. Nineteen percent of fathers had some high school education, 14% had a high school diploma or equivalent, 33% had some college, and 23% had a 4-year college degree or more. Fathers were 55% Latino, 31% White, 9% Black, and 5% other. Thirty-two percent were foreign born. Mothers were 39 years old on average ($SD = 7.24$, range 26-54), and 50% were not employed, 15% were part-time employed, and 35% full time employed. Thirty-six percent of mothers had some high school education, 20% had a high school diploma or equivalent, 20% had some college, and 23% had a 4-year college degree or more. Mothers were 65% Latina, 22% White, 6% Black, and 6% other, and 53% were foreign born. The ethnic and racial composition of the sample was reflective of the patient population at this Childrens Hospital as well as the surrounding urban area.

Measures

Received Social Support

Received social support was measured using the UCLA Social Support Interview (Wills & Shinar, 2000). Parents individually identified "the first names of the five most helpful people

during [their] son or daughter's treatment." Respondents could identify as few as zero helpful individuals or as many as five and could indicate anyone as a possible source of support (i.e., parents were not limited to choosing from a pre-existing list of possible sources of support). One-hundred and forty-nine social support sources were identified for 40 parents.

Respondents were only asked for the first name or initials of each alter, but most identified the source of support by full name and relationship (e.g., my mother, my aunt). To be conservative in identifying a shared source of support, when both names and relationships matched, the source was considered shared (e.g., mother's mother Maria and father's mother-in-law Maria). Non-matching members of parents' received support networks were considered unique. This technique may have underestimated the number of shared sources of support.

Type of Social Support Received

For each source of support identified, respondents were asked to identify the type and quality of support received. Using Wills and Shinar's (2000) definitions of instrumental, emotional, and informational support, descriptions of each type of support accompanied each item. Respondents were asked to identify the amount of instrumental support received with a single item, "How often did this person provide help by taking care of other children, offer transportation or money?" the amount of emotional support received with a single item, "How often did this person listen to your concerns or talk about how you were feeling?" and received information support with single item, "How often did this person provide information about health care or health insurance or types of cancer treatment?" All three items were measured on a five-point scale (0 = None, 4 = A lot). For each person identified, respondents were also asked to evaluate the overall quality of the support received from that person on a single semantic-differential scale (1 = Not Good, 7 = Very Good) (see Table 1).

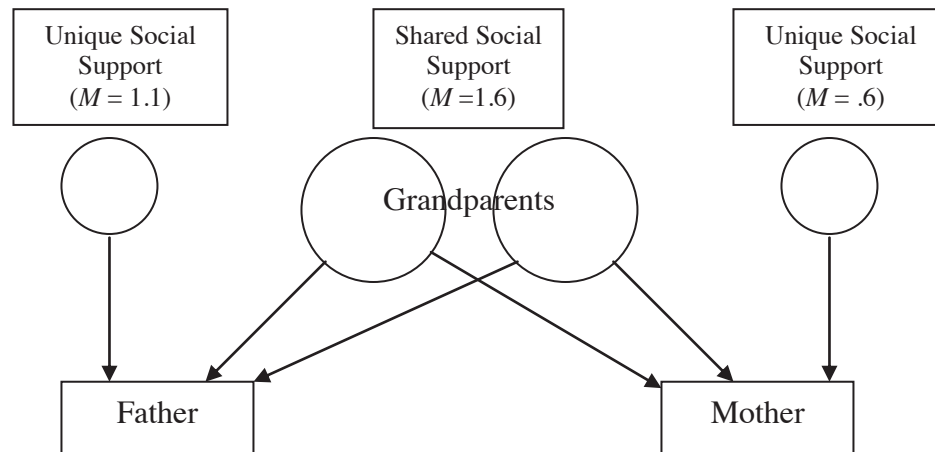
Using MLM, each of the social support sources were treated as Level 1 observations, and parents were treated as Level 2 predictors. This method controls for the dependence of the alters on the ego by treating alters as a consequence of the ego (for more MLM details see Kenny et al., 2006). The data were analyzed using LISREL 8.8 (Jöreskog, & Sörbom, 1996).

RESULTS

Analysis of support sources revealed that on average parents identified three people from whom they received social support ($M = 3.40$, $SD = 1.69$, $mdn = 4$, $mode = 5$, range 1 to 5). Most of this support came from family (63%), primarily participants' parents (i.e., the child's grandparents). Most sources of social support were women (69%), indicating that women were more often providers of social support. Less than half of social support members (1.60 of 3.40) were shared by both parents, and shared support sources were often family members. See Figure 1 for "average" family sociogram. Comparisons of fathers' and mothers' social support showed that parents received support from a similar number of sources ($M_{father} = 3.77$, $SD = 1.45$, $M_{mother} = 3.13$, $SD = 1.81$), $t(38) = 1.40$, $p = ns$.

To explore which characteristics of network alters predicts the overall quality of social support, the overall quality of the support was treated as a Level 1 dependent variable and the three types of social support (i.e., instrumental, emotional, informational), the sex of the source, the sex of the parent, whether the support was from a family member, and whether the support source was shared with the spouse were treated as Level 1 fixed effects. Multilevel modeling is concerned with model fit as well as the parameter estimates (Roberts, 2004). The best fitting model only included instrumental and emotional support amounts and family membership as fixed effects. The results indicated that amount of instrumental support ($\beta = .21$, $SE = .06$, $WALD = 3.76$, $p < .001$), emotional support ($\beta = .45$, $SE = .08$, $WALD = 5.50$, $p < .001$), and being a family member ($\beta = .20$, $SE = .11$, $WALD = 1.89$, $p < .05$) were

Figure 1. “Typical” Family Sociogram



related to overall support quality. To explore whether the overall quality of the source of social support was moderated by parent sex, several interaction terms were created: (1) being a shared source of support and sex of parent and (2) type of support (instrumental, emotional, health information) and sex of parent. None of the interaction terms were significant. For both mothers and fathers, more emotional and instrumental support received from family members increased overall support quality, and this relationship was not moderated by parent sex.

To answer questions regarding fathers’ inability to receive support, the differences between mothers and fathers in unique social support were explored using four paired samples *t* tests. Therefore, only parents’ unique sources of social support were included in this analysis (*N* = 68). In comparison to fathers (*M* = 2.08, *SD* = 1.07), mothers received more instrumental support from unique sources of social support (*M* = 2.57, *SD* = .93), *t*(66) = 2.00, *p* < .05, *d* = .49. From their unique sources of social support, the overall quality of social support for fathers (*M* = 6.40, *SD* = .90) was significantly less than the support mothers received (*M* = 6.80, *SD* = .37), *t*(66) = 2.17, *p* < .05, *d* = .58.

Table 1. All Means And Standard Deviations By Sex Of Respondent

	Mean		Standard Deviation		Sample Size	
	Men	Women	Men	Women	Men	Women
No. Sources of Support	3.77	3.13	1.45	1.81	20	20
Emotional Support	2.13	2.09	.85	1.08	20	20
Instrumental Support	2.36	2.39	.92	.97	20	20
Health Care Support	1.07	1.20	.97	1.14	20	20
Overall Social Support	6.23	6.11	1.17	1.80	20	20
% Sources Female	64	73			80	69
% Sources Family	63	63			80	69

DISCUSSION

Most researchers using egocentric network methods aggregate scores from multiple sources of support and do not explore the identities or unique qualities of these sources (Bissette et al., 2000). This case study demonstrated the value of treating each support source uniquely. By linking parents' received support networks into a common network for the couple all sources could be identified then as either shared--identified by both parents -- or unique--identified by only one parent. Linking the identities of network members to create a sociogram is a common technique in SNA (Wasserman & Faust, 1994) and past research has demonstrated the interdependence of spousal support networks (Widmer, 2006), but this case study demonstrates that linking couples' support networks increases the utility of egocentric data.

Sex Differences in Social Support

To answer the research questions in the case study, results suggest that sex differences in social support depend upon whether the network alter is shared or unique. Past research has documented fathers' lack of social support, both in quality and in quantity (Hoekstra-Weebers et al., 1999; Hoekstra-Weebers et al., 2000; Hoekstra-Weebers et al., 2001; Sloper, 2000). However, in this case study men and women received equivalent support in both quantity and quality. The results also failed to demonstrate the value of support provided by a female alter, but demonstrated the value of support provided by family members. In the context of childhood cancer, the familial relationship rather than the sex of the support source is associated with higher quality support. In addition, the interaction analyses demonstrated that it is not the case that shared sources of support are providing more support for mothers in comparison to fathers. Instead, emotional and instrumental support provided by family members is equally predictive of support quality for both parents. The differences between parents emerged only when separating unique from shared support sources. When considering

unique sources, mothers received more instrumental support of higher quality. The medium effect sizes suggest that the differences in support shown in past research may have resulted from unique, not shared sources of support. This offers support to Hoekstra-Weebers and colleagues' suggestion (2001) that men's lack of support may be a result of a less capable support network. The results of this study further refines this explanation by demonstrating that the lack of capability resides in the unique sources of support. Shared sources of support appear to be providing quality support to both mothers and fathers.

Directions, Applications, and Extensions

In addition to exploring social support, there are many applications of shared versus unique network alters. Egocentric data could be collected from many types of couples (e.g., friends, business colleagues, advisor-advisee). In addition to retaining the unique variance of each network alter, the methods described here can demonstrate how dyadic relationships affect network alters. For example, sharing a friend with a spouse may directly impact the friendship itself. Alters may be treated differently precisely because of their unique v. shared status in terms of the type, depth, and breadth of information shared. The consequences of making a unique source into a shared source (e.g., gaining in-laws) or developing a unique relationship with a previously shared source (e.g., during divorce) could provide some useful micro-level analyses of network level changes in connectedness. Other concepts, such as network density, may also be modeled as a Level 2 predictor. Although this study did not explore the interconnections between the sources of social support, past research has advocated asking whether and how network alters are associated to one another (Burt, 1984). In the case of social support for families, if shared sources of support are tightly connected, they may diffuse responsibility and support the family more effectively (Street, 2003). If sources of support are disconnected, supporters might not be able to

accurately assess what support is needed or provide targeted support.

Researchers using traditional SNA, but working with a large amount of missing data, may also find value in MLM. This would likely be most attractive for researchers who have collected information beyond network ties, such as attributes about the ego and evaluations of the alters or ties by the ego. If overlap existed among ego-networks, shared and unique ties might also be identified.

The methods used in this case study can be improved in future investigations. Although asking dyad members to complete surveys in isolation from each other helps to reduce dyad members' influence on one another during data collection, once surveys are complete, researchers may want to ask dyads to identify their network alters as shared or unique. Researcher-matched alters, such as those in this case study, may be subject to errors of identification, especially if few or no details are offered about an alter (e.g., only initials, missing data). Furthermore, future work should increase the sample size to increase the possibility of exploring interaction effects and improve the ability to detect small effect sizes.

Hopefully, the methods and procedures described here will provide some useful guidance and be a source of fruitful application for researchers using egocentric data from couples. As interest in dyadic data grows, social network researchers can provide key insights into the ways couples behave within their social environments.

REFERENCES

- Brissette, I., Cohen, S., & Seeman, T. E. (2000). Measuring social integration and social networks. In S. Cohen, L. G. Underwood, & B. H. Gottlieb (Eds.), *Social support measurement and intervention: A guide for health and social scientists* (pp. 53-85). Oxford: University Press.
- Burleson, B. R., & Kunkel, A. (2006). Revisiting the different cultures thesis: An assessment of sex differences and similarities in supportive communication. In K. Dindia & D. J. Canary (Eds.), *Sex differences and similarities in communication* (2nd ed.) (pp. 117-159). Mahwah, NJ: Erlbaum.
- Burt, R. S. (1984). Network items and the General Social Survey. *Social Networks*, 6, 93-339.
- Cutrona, C. E. (1996). *Social support in couples*. Thousand Oaks, CA: Sage.
- Hoekstra-Weebers, J. E. H. M., Jaspers, J. P. C., Kamps, W. A., & Klip, E. C. (1998). Gender differences in psychological adaptation and coping in parents of pediatric cancer patients. *Psychooncology*, 7, 26-36.
- Hoekstra-Weebers, J. E. H. M., Jaspers, J. P. C., Kamps, W. A., & Klip, E. C. (1999). Risk factors for psychological maladjustment of parents with children with cancer. *Journal of The American Academy of Child and Adolescent Psychiatry*, 38, 1526-35.
- Hoekstra-Weebers, J. E. H. M., Jaspers, J. P. C., & Kamps, W. A. (2000). Factors contributing to the psychological adjustment of parents of paediatric cancer patients. In L. Baider, C.L. Cooper, & A. Kaplan De-Nour (Eds.), *Cancer and the family* 2nd ed. (pp. 73-92). Chichester, UK: John Wiley & Sons.
- Hoekstra-Weebers, J. E. H. M., Jaspers, J. P. C., Kamps, W. A., & Klip, E. C. (2001). Psychological adaptation and social support of parents of pediatric cancer patients: A prospective longitudinal study. *Journal of Pediatric Psychology*, 26, 225-235.
- Jöreskog, K., & Sörbom, D. (1996). *LISREL 8: User's reference guide*. Lincolnwood, IL: Scientific Software International.
- Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.
- Marsden, P. V. (1987). Core discussion networks of Americans. *American Sociological Review*, 52, 122-131.
- Marsden, P. V. (1990). Network data and measurement. *Annual Review of Sociology*, 16, 435-63.
- Roberts, J. K. (2004). An introductory primer on multilevel and hierarchical linear modeling. *Learning Disabilities: A Contemporary Journal*, 2, 30-38.

- Scott, J. (1991). *Social network analysis*. London: Sage.
- Segrin, C., & Flora, J. (2005). *Family communication*. Mahwah, NJ: Erlbaum.
- Sloper, P. (2000). Predictors of distress in parents of children with cancer: A prospective study. *Journal of Pediatric Psychology, 25*, 79-91.
- Stevens, N., & Westerhof, G. J. (2006). Partners and others: Social provisions and loneliness among married Dutch men and women in the second half of life. *Journal of Social and Personal Relationships, 23*, 921-941.
- Street, R. L. (2003). Communication in medical encounters: An ecological perspective. In T. L. Thompson, A. M. Dorsey, K. I. Miller, & R. Parrott (Eds.), *Handbook of Health Communication* (pp. 63-89). Mahwah, NJ: Erlbaum.
- Thoits, P. A. (1995). Stress, coping, and social support processes: Where are we? What next? *Journal of Health and Social Behavior, (extra issue)*, 53-79.
- Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.
- Widmer, E. D. (2006). Who are my family members? Bridging and binding social capital in family configurations. *Journal of Social and Personal Relationships, 23*, 979-998.
- Wills, T. A., & Shinar, O. (2000). Measuring perceived and received social support. In S. Cohen, L. G. Underwood, & B. H. Gottlieb (Eds.), *Social support measurement and intervention: A guide for health and social scientists* (pp. 86-112). Oxford: University Press.

What You Believe Travels Differently: Information and Infection Dynamics Across Sub-Networks

Patrick Grim

*Department of Philosophy
Stony Brook University*

Christopher Reade

*Gerald R. Ford School of Public Policy
University of Michigan*

Daniel J. Singer

*Department of Philosophy
University of Michigan*

Steven Fisher

*Center for Study of Complex Systems
University of Michigan*

Stephen Majewicz

*Department of Mathematics
Kingsborough Community College*

Abstract

In order to understand the transmission of a disease across a population we will have to understand not only the dynamics of contact infection but the transfer of health-care beliefs and resulting health-care behaviors across that population. This paper is a first step in that direction, focusing on the contrasting role of linkage or isolation between sub-networks in (a) contact infection and (b) belief transfer. Using both analytical tools and agent-based simulations we show that it is the structure of a network that is primary for predicting contact infection—whether the networks or sub-networks at issue are distributed ring networks or total networks (hubs, wheels, small world, random, or scale-free for example). Measured in terms of time to total infection, degree of linkage between sub-networks plays a minor role. The case of belief is importantly different. Using a simplified model of belief reinforcement, and measuring belief transfer in terms of time to community consensus, we show that degree of linkage between sub-networks plays a major role in social communication of beliefs. Here, in contrast to the case of contract infection, network type turns out to be of relatively minor importance. What you believe travels differently. In a final section we show that the pattern of belief transfer exhibits a classic power law regardless of the type of network involved.

Authors

Patrick Grim is Distinguished Teaching Professor of Philosophy at Stony Brook, co-author of *The Philosophical Computer* (MIT 1998) and *Beyond Sets* (Ontos/Verlag 2011). Grim's publications appear not only in philosophy but in computer science, theoretical biology, cognitive science, linguistics, and game theory.

Christopher Reade is a consultant at ThoughtWorks Inc. and a graduate of the Gerald R. Ford School of Public Policy at the University of Michigan.

Steven Fisher is an undergraduate in Complex Systems studying networks and network theory in the Center for the Study of Complex Systems at the University of Michigan.

Daniel J. Singer is a Ph.D. candidate in Philosophy at the University of Michigan working on foundational questions of epistemology and formal epistemology.

Dr. Stephen Majewicz is an Associate Professor in Mathematics at Kingsborough Community College. His fields of interest include nilpotent groups, exponential A-groups and nilpotent R-powered groups, combinatorial group theory, and axiomatic set theory.

Acknowledgements

This work was supported in part by the National Institute of General Medical Sciences MIDAS grant 1U54GM088491-01, Computational Models of Infectious Disease Threats, and by a pilot grant for Developing an Agent-Based Model to Assess Racial Differences in Medical Discrimination, Social Support and Trust, administered by the Graduate School for Public Health at the University of Pittsburgh.

Please contact Patrick Grim for correspondence at pgrim@notes.cc.sunysb.edu

INTRODUCTION

Public health has been a primary target for agent-based and network modeling. A significant amount of work has been done on the role of network structure in the spread of disease (Meyers, Pourbohloul, Newman, Skowronski & Brunham 2005; Keeling 2005; Ferrari, Bansal, Meyers & Bjørnstad 2006; Miller & Hyman 2007; Eubank, Guclu, Kumar, Marathe, Srinivasan, Toroczkai & Wang 2004). But it is clear that health-care behaviors are as crucial in the pattern of any pandemic as are the biological characteristics of the pathogens involved (Epstein, Parker, Cummings & Hammond 2008; Auld 2003; Del Valle, Hethcote, Hyman, & Castillo-Chavez 2005; Barrett, Bisset, Leidig, Marathe, & Marathe 2009; Funk, Gilad, Watkins, & Jansen 2009; Hallett, Gregson, Lewis, Lopman, & Garnett 2007). Those health-care behaviors are contingent on beliefs. On standard models, these include at least beliefs regarding severity, susceptibility, effectiveness and the cost of preventive measures (Harrison, Mullen, & Green 1992; Janz & Becker, 1984; Mullen, Hersey, & Iverson 1987; Strecher & Rosenstock 1997).

In order to understand the spread of disease we will have to better understand the spread of beliefs and behaviors. Moreover, as public health interventions are often targeted to beliefs and behaviors we will have to better understand the spread of beliefs and behaviors in order to intervene effectively. For a better picture of disease dynamics and to better the prospects for effective intervention we need a better understanding of the dynamics of belief transmission across social networks. Although important empirical work has been done on social networks and the diffusion of beliefs and behaviors (Valente 1995, 2010; Morris, Podhisita, Wawer & Handcock 1996; Morris 1997; Valente & Davis, 1999; Kincaid 2000; Hamilton, Handcock & Morris 2008), significantly less has been done with the tools of agent-based modeling toward understanding the abstract dynamics of belief (see however

Centola & Macy 2007 and Golub & Jackson, forthcoming).¹

In what follows we take some steps in that direction, with an emphasis on the pervasive social phenomenon of sub-network groups or clusters. Our social networks do not form a uniform and homogenous web. Social communities are composed of sub-communities, with varying degrees of contact and isolation between them; both in terms of the physical contact necessary for disease transmission and the informational contact crucial to the transmission of belief. Racial, ethnic, socio-economic, demographic, and geographical sub-communities offer a clear example. Racial and economic sub-communities may be more or less isolated or integrated with other sub-communities, with varying strengths of information transfer, communication, and trust. In the case of a pandemic, degree of isolation or integration will be crucial in predicting the course of contact and therefore the dynamics of disease transmission. But in such a case degree of informational isolation or integration will also be crucial in tracking changes in health care beliefs and behaviors, with both immediate and long-range effects on the course of the disease.

What we offer is an abstract model of this very real phenomenon. We track the role of degree of linkage between sub-networks in the transfer of disease and the transfer of information, with contrasting results in the two cases. Linkages between sub-networks have also been termed 'bridges,' analogous to a concept of bridges in computer networking and identified in Trotter,

¹ Centola and May consider 'complex contagions', in which more than one neighbor is required for infection. This is not strictly speaking a reinforcement effect, but does show dynamics similar to that studied for belief reinforcement here—and a similar contrast with simple infection. Golub and Jackson outline analytic results on 'homophily' in random networks, with a similar emphasis on the contrast between diffusion and belief averaging. Our work here, part analytic and part from agent-based simulations, extends that work and shows that the central contrast holds across networks of various types.

Rothenberg and Coyle (1995) as a key area for future work in network studies and health care. L. C. Freeman (1977) speaks of degree of linkage in terms of segregation and integration between sub-networks. Ours is a formal study of networks, however, and such a terminology may carry distracting connotations. Homophilous networks, in which nodes link preferentially with others with similar characteristics, often take the form of clustered sub-networks with limited degrees of linkage; precisely the type we study here. Our focus is on the implications of a network structure, however, not how a network may have acquired that structure.

We focus on the structure of contact and informational networks and the impact of that structure on the dynamics of infection and information. In the first section we outline simple analytic results and a wider spread of agent-based simulation results regarding the impact of degree of linkage between sub-networks on the spread of infection across a community. Those results regarding simple diffusion serve as a base of comparison for the very different results regarding the effects of degree of linkage on the transmission of beliefs.

The dynamics of belief turns out to be very different from the dynamics of contact infection. For infection, measured in terms of average time to total infection across a network, it is the structure of the network or its sub-networks that is of primary importance—whether the basic network or networks at issue form rings, total networks, hubs, wheels, small worlds, scale-free or random networks. The degree of linkage between sub-networks of such a type is of relatively minor importance for infection. For belief transmission on the model we construct, in contrast, measured in terms of average time to total consensus, network structure is of minor significance. Where the dynamics of belief is at issue, it is the degree of linkage between sub-networks that is of primary importance. The effect of degree of linkage on belief change, we show, regardless of network type, shows the pattern of a classic power law.

Our effort here is to emphasize a basic point regarding the different dynamics of belief and infection across networks. More complete details of both analytic results and results from simulation are available in an on-line appendix at www.pgrim.org/connections.

Infection Dynamics Across Linked Sub-Networks

Ring and Total Network: A First Example

Figure 1 shows a series of four network structures, clearly related in terms of structure. The network on the left is a single total network, also known as a complete network or maximal graph. The three pairs on the right form paired sub-networks with increasing numbers of connecting links. We will use degree of linkage in a relative sense to refer to increased connecting links or bridges of this sort. A quantitative measure is possible in terms of the number of actual linkages between nodes of distinct groups or sub-networks over the total possible.²

We focus on varying degrees of connection between sub-networks of varying structure. For simplicity we use just two sub-networks of equal size, concentrating on ring sub-networks, total or connected networks, small worlds, random and scale-free sub-networks. How does the degree of connection between two sub-networks affect the dynamics of diffusion or infection across the network as a whole? How do results on degree of connection between sub-networks of a specific structure compare with results on a single network of the same structure to which the same number of links are added? Here theoretical fundamentals trace to Granovetter 1973; and an early example of

² Full linkage between total sub-networks, such that every node in one sub-network will connect to every node in the other sub-network, will result in the single total network on the left. But of course it will not hold in general that full linkage between sub-networks of type x will result in a single network of type x : full linkage between ring networks will not result in a single ring.

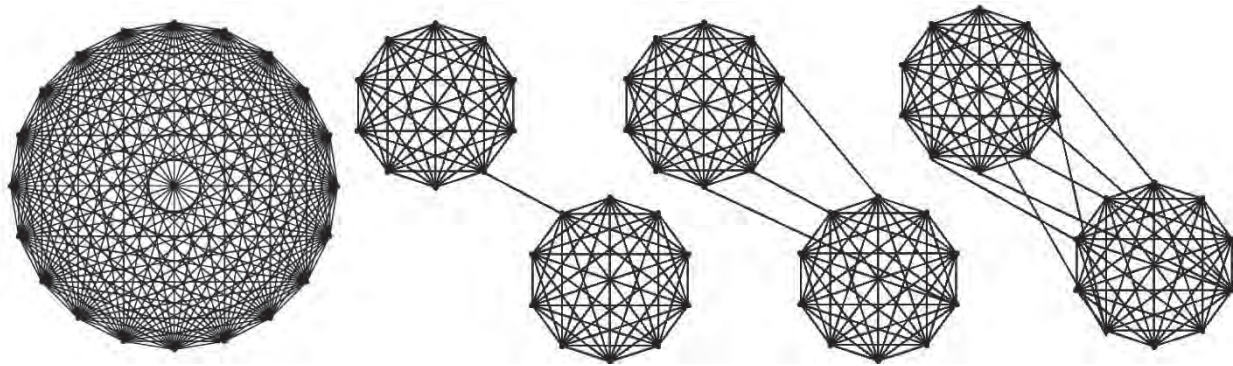


Figure 1. A Single Total Network and Increased Degrees of Linkage Between Total Sub-networks

network analysis regarding infection appears in Klondahl 1985.

Some results are simple and analytic, but also indicate the variety that can be expected. Consider, at one extreme, a network composed of two totally connected sub-networks with a single link between them, as in the second network in Figure 1. How many steps will be required to total infection, starting from a single random infected node? Assuming a 100% infection rate, where n is the total number of nodes, the average number of steps to total infection is:

$$\frac{3(n - 2) + 4}{n} = \frac{(3n - 2)}{n} .$$

where n is the total number of nodes. From any node other than those on the ends of our connecting link, there are three steps to total infection: (1) to all nodes of the immediate connected networks, (2) across the one connecting link, and (3) from there to all nodes of the opposite connected network. If the initially infected node is one of those on the ends of our connecting link, there are merely two steps to total infection, giving us the formula above.

Adding further links has no dramatic effect in such a case. Because our sub-networks are totally connected, a first step in every case

infects all nodes in a sub-network; from there any number of links between sub-networks merely transfer the infection to the second sub-network. For a network with two sub-networks of equal size, therefore, again assuming an infection rate of 100% rate and incorporating n nodes and m discrete links between sub-networks (links sharing no nodes),³ the average time to total infection will be simply:

$$\frac{3(n - 2m) + 4m}{n} = \frac{(3n - 2m)}{n} .$$

As n increases relative to $m \neq 0$, time to infection approaches a limit of 3. As m increases relative to n , with a limit of $m = .5 n$, time to infection approaches a limit of 2. For a single total network, like that on the left in Figure 1, any 'added' linkages would simply be redundant, with no effect at all: infection will in all cases be in a single step.

Where sub-networks are total, variance in infection time is necessarily just between 2 and 3 steps. At the other extreme is the case of a network with rings as sub-components. Here variance in infection time is much greater. The maximal number of steps to full infection from a single node across a ring sub-network is $s/2$

³ In order to keep the outline of basic relationships as simple as possible we ignore the complication that links can share a single node at one end.

with s as the number of nodes for that sub-network where s is even, or $(s - 1)/2$ in the case of odd numbers of nodes. The longest time for diffusion across a network of two equal-sized rings each with an even number of nodes $n/2$ is therefore:

$$\frac{n}{4} + 1 + \frac{n}{4} = \frac{n}{2} + 1 .$$

Where the number of nodes $n/2$ in each sub-network is odd the maximal number of steps is:

$$\frac{\frac{n}{2} - 1}{2} + 1 + \frac{\frac{n}{2} - 1}{2} = \frac{n}{2} .$$

If the source of infection is one of the nodes on the end of a bridge between sub-networks, time to infection will be minimal: where $n/2$ is even the minimal time to infection will be $\frac{n}{4} + 1$; where $n/2$ is odd, time to infection will be $\frac{n}{4} + \frac{1}{2}$.

Variance between maximum and minimum times to total infection is therefore extremely sensitive to the structure of sub-networks. In the case of total sub-networks, that variance is simply 1 regardless of the number of nodes. In the case of ring sub-networks, the variance is close to $n/4$. The consequences for prediction are clear: to the extent that a social network approaches a total network, point predictions of infection times can be made with a high degree of confidence. To the extent that a social network approaches a ring, on the other hand, point predictions will not be possible without wide qualification.

The structure of sub-networks is crucial for other factors as well. We have noted that increasing links between sub-networks has a minimal effect where those sub-networks are total. Where sub-networks are rings of 50 nodes, in contrast, the effect is dramatic. The

top line in Figure 2 shows results from a computer-instantiated agent-based model in which we progressively increase the number of links between random nodes of those sub-networks from 1 to 50. For each number between 1 and 50 we create 1000 networks with random links of that number between sub-networks, taking the average over the 1000 runs. For ring sub-networks the time to full infection decreases from an average of 38.1 steps for cases in which there is a single link between ring sub-networks to 7.6 for cases in which there are 50 links.

Similar simulation results for added links between total sub-networks, in contrast, show a relatively flat result with decline in average time to infection from only 2.98 to 2.35. Difference in network structure clearly makes a major difference in time to total infection. That difference is not due to degree of linkage between sub-networks, however. A graph of results in which links are added across a single ring and not between ring sub-networks shows a result almost identical to that in Figure 2.

The lesson from ring and total networks is that it is not the degree of linkage between sub-networks that affects time to total infection but overall network structure itself, whether characterizing a single network or linked sub-networks. Changes in infection rates with additional random links (1) across a single network and (2) between two smaller networks with the same structure show very much the same pattern. Degrees of linkage between sub-networks interact with the structure of those sub-networks in order to generate patterns of infection, but it is the structure of the networks rather than the degree of linkage that plays the primary role. Analytical and simulation results for hub and wheel networks, very much in line with conclusions above, are available in an online appendix (www.pgrim.org/connections).

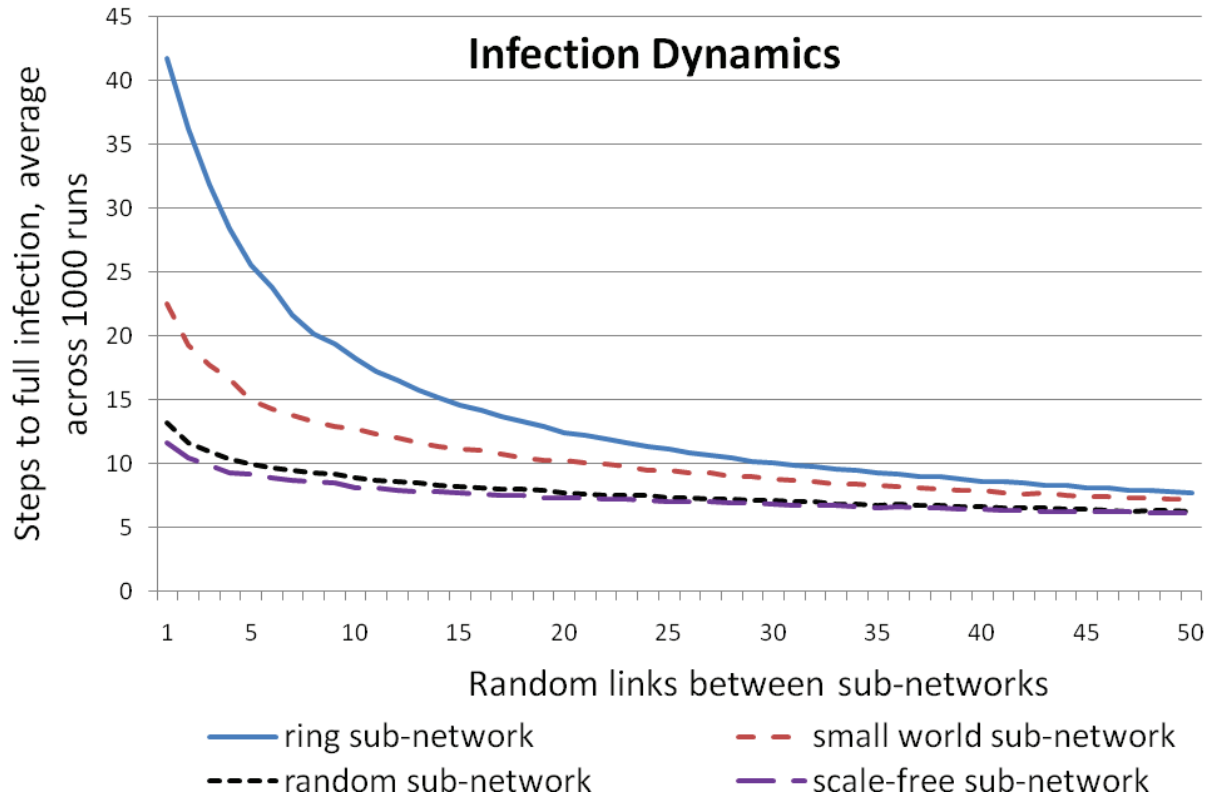


Figure 2. Average Time to Total Infection with Increasing Links Between Sub-Networks

Infection Across Small World, Random, and Scale-Free Networks

For patterns of infection, the importance of general structure type over degree of linkage between sub-networks holds for small world, scale free, and random networks as well. Results for small world networks are shown in the second line from the top in Figure 2 with roughly a 9% probability of rewiring for each node in an initial single ring (see Watts & Strogatz 1998).⁴ Increasing linkages between sub-networks from 1 to 50 results in a decrease in steps to total infection from 22.5 steps to 7.45. Increasing links within a single small

world follows virtually the same pattern, with a decrease from 19.8 to 7.2.

Similar results for random and scale-free networks appear in the third and fourth graphed lines of Figure 4. For random networks, roughly 4.5 percent of possible connections are instantiated within each sub-network, with minimal links needed to guarantee connected networks. Our scale-free networks are constructed by the preferential attachment algorithm of Barabási and Albert (1999).

Here as before there is little difference where additional links are added within a single network, whether small-world or scale-free. In each case the number of initial steps is slightly smaller, but only in the first 10 steps or so is there any significant difference and convergence is to the same point. In the case of random networks, time decreases from 9.79 to 6.45. In

⁴ Our probability is 'roughly' 9% because in each case we add minimal links so as to assure a connected network. Without that assurance, of course, infection is not guaranteed to percolate through the network as a whole.

the case of scale-free networks, time decreases from 7.9 to 6.08.

In all the cases considered, it is not degree of linkage between sub-networks but the network structure involved in both single and linked sub-networks that produces network-specific signatures for infection. This largely accords with analytic results by Golub and Jackson (forthcoming) on diffusion dynamics across linked random networks.⁵ Golub and Jackson find that in the limit degree of linkage between random networks has no effect on time to total infection. What our results indicate is that such a result is by no means restricted to random networks, holding across network types quite generally. Where infection is concerned, a prediction of time to total infection demands a knowledge of the general structure of the contact network at issue—ring or total, for example, scale-free or random, but does not demand that we know whether it is a single network or a linked set of smaller networks of that same structure that is at issue.

Infection on Networks: Qualifications and Provisos

Results to this point have been calculated with an assumption of 100% infection—a disease guaranteed to be transmitted at every time-point of contact between individuals. More realistic assumptions regarding rate of infection affect the rates calculated above, more pointedly emphasizing the importance of structure. Here we again use ring and total networks as an example.

Where sub-networks are total, probability of infection from single contact really makes a difference only at the link between sub-networks: as long as the probability of infection exceeds $2/n$, a quick infection of all individuals

in the total sub-networks is virtually guaranteed. Simulation results indicate that with a single link between total sub-networks the average time to full infection shifts only from an average of 3.8 steps to an average of 2.98 with a change of infection rate from 100% to 50%. For ring sub-networks, on the other hand, the same change in infection rate roughly doubles the time to full infection across all numbers of linkages.

For more realistic infection rates, therefore, it is more important rather than less to know the structure of social networks. If those sub-networks approximate total networks, neither infection rate nor additional links between sub-networks make much difference. If sub-networks approximate ring networks, both number of links and infection rate will make a dramatic difference in the course of an infection.

Where average time to infection is our measure, degree of linkage between sub-networks as opposed to additional links within a single network of that structure is not of particular significance. But here we need to add an important proviso: this does not mean that the course of an epidemic across a single network and across sub-networks with various degrees of linkage is not significantly different. That dynamic is often very different—in ways that might be important for intervention, for example—even where average time to total infection is the same. The typical graphs in Figure 3 show the rate of new infections over time for (a) a single network and (b) linked sub-networks of that type. Single networks show a smooth normal curve of increasing and declining rates of new infection. Linked sub-networks show a saddle of slower infection between two more rapid peaks.

Despite uniformity of predicted time to total infection, therefore, sparsely linked sub-networks will always be 'fragile' at those links, with temporal saddle points in the course of an epidemic to match. Those weak linkages and saddle points offer crucial opportunities for targeted vaccination in advance of an epidemic, or intervention in the course of it.

⁵ Golub and Jackson characterize their results using the term 'homophily', defined in terms of the relative probability of node connection within as opposed to outside of a group or sub-network. For random networks, though not for other network structures, this corresponds to the degree of linkage between sub-networks that is our focus here.

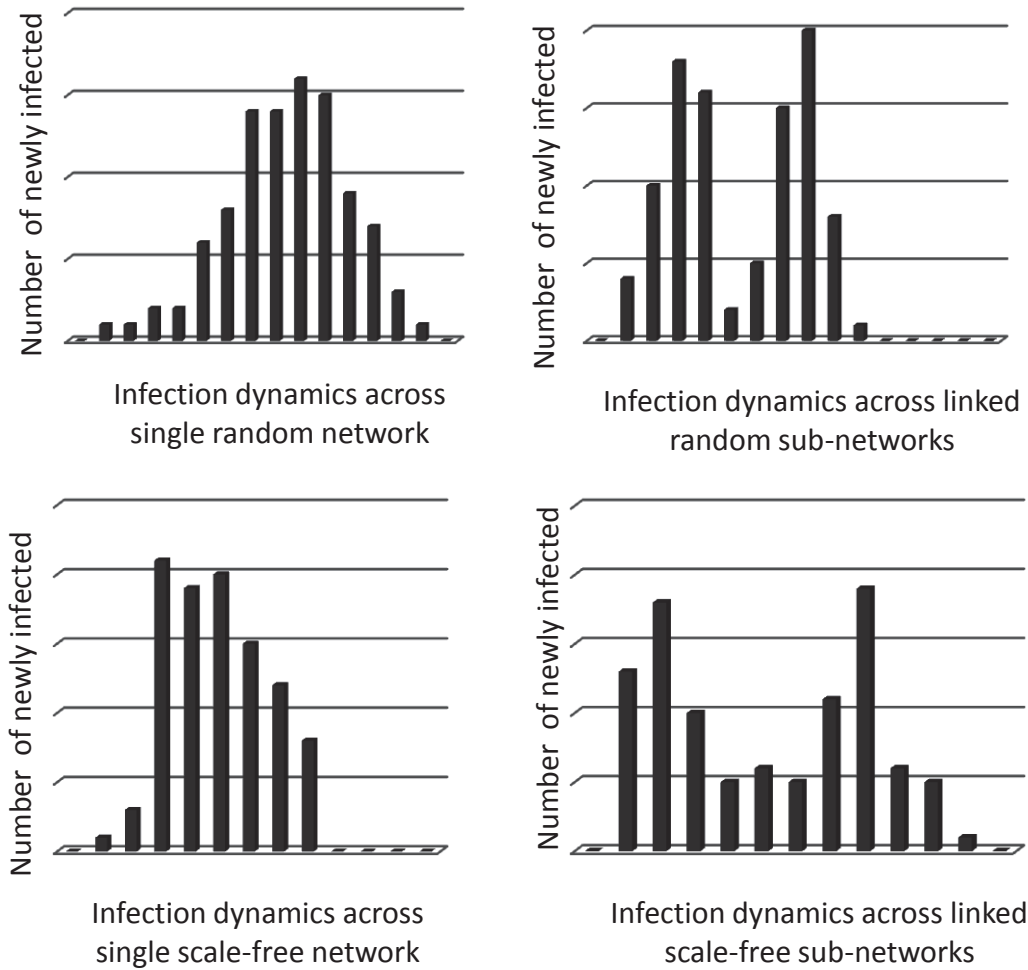


Figure 3. Contrasting Dynamics of Infection in Single and Linked Sub-Networks

Information Dynamics Across Linked Sub-Networks

What you believe travels differently. In what follows we use a simple model of belief updating to show the crucial importance of degree of sub-network linkage in belief or information transmission across a network. Some earlier results have noted similarities in infection dynamics and the spread of ideas (Newman 2001, Redner 1998, Börner et. al. 2003). Our purpose is to emphasize crucial differences between them.

In this first model our agents' beliefs are represented as a single number between 0 and 1. These are beliefs in the severity of a disease, perhaps, the probability of contracting the disease, or the effectiveness of vaccination. (Harrison, Mullen, & Green 1992; Janz & Becker, 1984; Mullen, Hersey, and Iverson, 1987; Strecher & Rosenstock, 1997). Agents are influenced by the beliefs of those around them, updating their belief representation in terms of the beliefs of those with whom they have information linkages.

To this extent we can argue that the model is relatively realistic: some beliefs can be

represented on such a scale, and people are influenced to change those beliefs by, among other things, the expressed beliefs of those with whom they have contact. What is admittedly unrealistic is the simple form of belief updating we use in the model: an averaging of current beliefs with those with whom one has network contact. No-one thinks that averaging of beliefs in an informational neighborhood captures the real dynamics of belief change. Such a mechanism does, however, instantiate a pattern of reinforcement: the more one's beliefs are like those of one's network neighbors, and the more they are like more of one's network neighbors, the less inclination there will be to change those beliefs. The more one's beliefs are out of sync with one's neighbors, the greater the pressure there will be to change one's beliefs.

That beliefs will change in accord with some pattern of reinforcement along those lines is very plausible, backed by a range of social psychological data, and is therefore an aspect of realism in the model. What is unrealistic is the particular form of reinforcement instantiated here—the particularly simple pattern of belief averaging, applied homogeneously across all agents. In order to be informative regarding an exterior reality, a model, like any theory, must capture relevant aspects of that reality. In order to offer both tractability and understanding, a model, like any theory, must simplify. This first model of belief transmission is intended to capture a reality of belief reinforcement; the admittedly artificial assumption of belief averaging is our simplification.⁶

Our attempt, then, is not to reproduce any particular pattern of realistic belief change but to emphasize the impact of certain predictable characteristics of belief change—with reinforcement a primary component—on the dynamics of belief. In particular, we want to emphasize the major differences between the dynamics of belief change across information

⁶ For background on both the importance and limit of realism in different forms of models, see Grim, Rosenberg, Rosenfeld, Anderson, & Eason 2010 and Rosenberg, Grim, Rosenfeld, Anderson & Eason 2010.

networks and the dynamics of infection diffusion across contact networks, outlined above. What you believe travels differently.

Given belief averaging, and regardless of initial assignment of belief representations, all agents in this model eventually approach the same belief value. We can therefore measure the effect of network structure on belief convergence by measuring the number of steps required on average until all agents in the network are within, say, a range of .1 above or below the mean belief across the network as a whole. In what follows we use this range of variance from the mean as our measure of convergence, averaging over 100 runs in each case.

We begin with polarized agents. Half of our agents are drawn from a pool with belief measures that form a normal distribution around .25, with a deviation of .06. The other half are drawn from a pool with belief measures in similar normal distribution around .75. In studying linked sub-networks our agents in one sub-network are drawn from the .25 pool; those in the other are drawn from the .75 pool. In the case of single networks agents are drawn randomly from each pool. We found belief polarization of this form to be necessary in order to study the effects of sub-network linkage in particular; were beliefs of all our agents merely randomized, convergence to an approximate mean could be expected to occur in each sub-network independently, and time to consensus would not then be an adequate measure of the effect of sub-network linkage.

Belief Diffusion across Ring and Total Networks

In outlining the dynamics of infection we contrasted linked sub-networks of particular structures—ring, small world, random, total, and scale-free—with single networks of the same structure. In exploring the dynamics of belief we will again study these types side by side. As we add additional links between sub-networks, how do the dynamics of belief diffusion change, measured in terms of time to consensus across the community.

We progressively add random links (1) between belief-polarized ring sub-networks, and (2) within a single ring network of belief-polarized agents. Average times to consensus are shown in Figure 4.

Increasing linkages between polarized ring sub-networks makes a dramatic difference. Average time to consensus for a single linkage in such a case is 692.44. The average time to consensus for 50 linkages is 11.59, with a distinct and characteristic curve between them. For infection, we noted, there is virtually no difference between added links within a single ring network and added links between ring sub-networks. In the case of belief, in contrast, there is a dramatic difference between the two graphs.

Within a single total network, all agents will achieve a mean belief in a single step; additional linkages in such a case are merely redundant. Results in total sub-networks, in contrast, parallel those for rings above. Average steps to belief convergence with a single link approximate 700 steps in both cases; with 50 links, average time to convergence is 12 in the case of rings and 16 in the case of total sub-networks. The overall pattern of the two graphs is also very much the same. What that similarity shows is the striking effect of degree linkage in each case: an effect that in the transmission of belief overrides the fact that we are dealing with totally distributed ring networks in one case, totally connected networks in the other.

Belief Transmission across Small World, Random, and Scale-Free Networks

The same contrasts between single and linked sub-networks in the case of belief transmission hold for other network structures as well. The effect of added linkages within a single small-world network closely parallels that for

the single ring shown above. Results for added linkages in small-world sub-networks are dramatically different. In absolute terms the results for small worlds differ from those shown for rings, declining from 481 steps to 11.4. The shape of the curve for small worlds, however, is very much that shown for rings above.

Given a single random network, using 2.25% of possible linkages, additional linkages give a decline in time to belief consensus from only approximately 6 steps to 4. Where random sub-networks are at issue (using 4.5% of possible linkages in each sub-network), the curve is again that displayed for rings above, though here absolute values decline from 244 to 10.15.

For single scale-free networks, additional linkages give a roughly linear decline from 20 to 7 steps. For scale-free sub-networks, additional linkages again follow the curve shown above, here with absolute values dipping from 325 to 11.73.

A similar curve characterizes effects of degree linkage in belief transmission regardless of the basic structure of the sub-networks involved. Although absolute values across that curve differ significantly, the shape of the curve does not. We emphasize this point in Figure 5 by plotting belief transmission results for sub-network types in log-log form.

Linkage degree effects follow the same pattern regardless of the structure of sub-networks. If one wants to plot the course of an epidemic, we noted in section I, it is crucial that one knows the structure of the networks involved. If one wants to plot the course of belief transmission, knowledge of structure is much less important.

The particular structure of networks is important in order to gauge whether a single link between sub-networks will allow consensus in 140 steps or 700, as indicated for hub and total networks

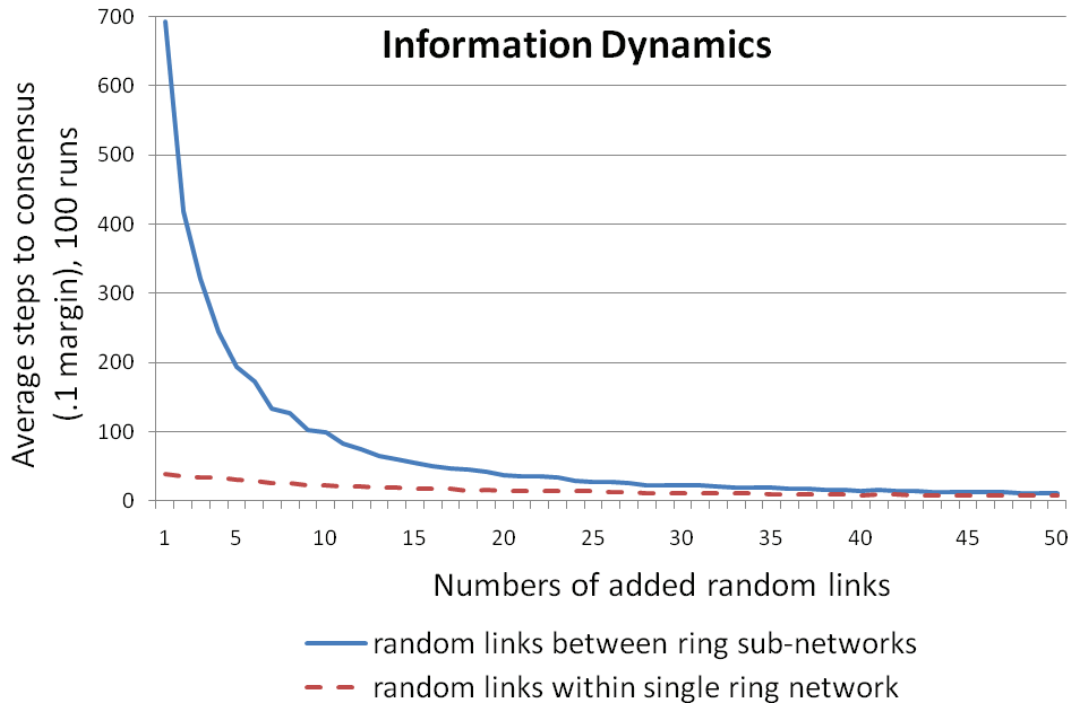


Figure 4. Time to Belief Consensus with Increasing Linkages in Single Ring and Between Ring Sub-Networks

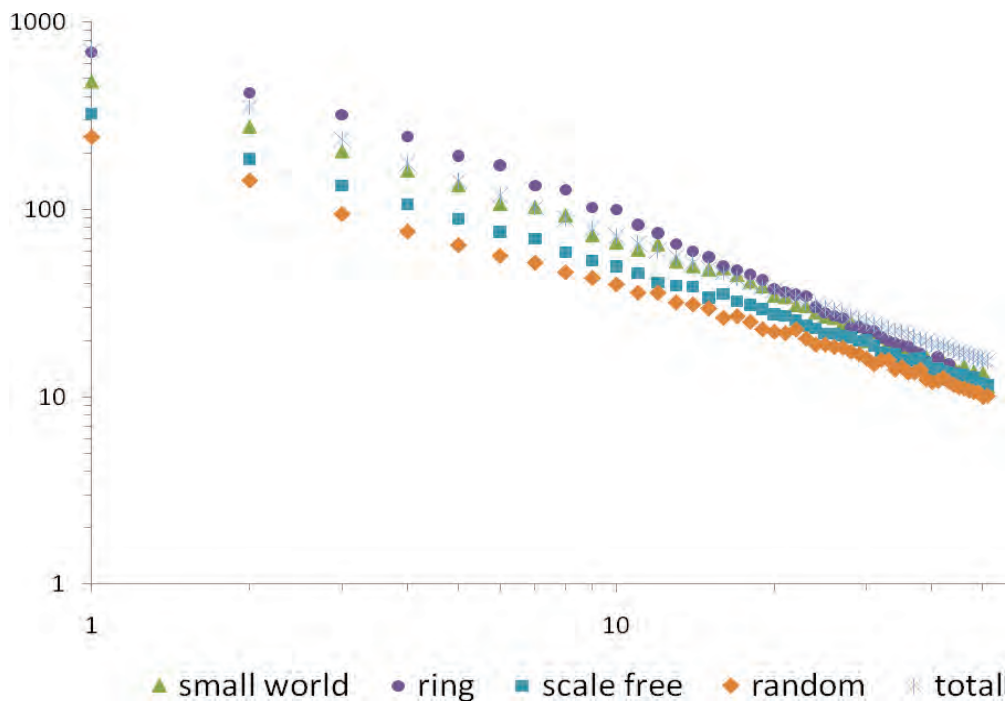


Figure 5. Time to Belief Consensus with Increasing Linkages Between Sub-Networks (plotted log-log)

in Figure 5. The pattern of changes in belief transmission with increasing linkages between sub-networks from any initial point, however, is precisely the same regardless of network structure. That pattern is the classic signature of power law distributions, indicating that the relationship between increased linkage and time to consensus parallels a range of natural and social phenomena, including the relationship between frequency and size of earthquakes, metabolic rate and body mass of a species, size of a city and the number of patents it produces. Power law distributions also appear in some empirically observed characteristics of biochemical, protein, citation and sexual contact networks (Faloutsos, Faloutsos, & Faloutsos, 1999; Jeong, Tombor, Albert, Ottvai, & Barbási 2000; Fell & Wagner 2000; Liljeros, Edling, Amaral, Stanley, & Åberg 2001; Newman 2001, 2005). The fact that such an effect appears in linkage effects on the dynamics of belief suggests the possibility of incorporating a range of theoretical and methodological work from other disciplines in studying behavior dynamics in the spread of disease, particularly with an eye to the effect of belief polarization, health care disparities, and social linkage or integration between ethnic and socio-economic sub-communities.

CONCLUSIONS & FUTURE WORK

Our focus here has been on the structure of contact and informational networks and the very different impact of aspects of that structure on the dynamics of infection and information.

For infection, measured in terms of average time to total infection across a network, it is the structure of the network or sub-networks that trumps other effects. In attempting to gauge time to total infection across a community, the primary piece of information needed is whether the social network or component networks at issue approximate rings, hubs, wheels, small worlds, random, scale-free or total networks. For time to total infection, degree of linkage between sub-networks is of much less importance, though we have noted that points of linkage continue to play an important role

with regard to fragility and prospects for targeted intervention.

For information, measured in terms of average time to belief consensus, the importance of general structure and linkage between sub-networks are reversed. On the model of belief used here, in attempting to gauge the dynamics of information flow across a community, the primary piece of information needed is the degree of linkage between composite sub-communities, whatever their internal structure. The fact that the particular structure of those sub-communities is of lesser importance is highlighted by the fact that average time to belief consensus given increasing linkages follows the same familiar power-law pattern regardless of networks structures involved.

It is quite plausible that belief transmission involves strong reinforcement effects; the model of belief used here is designed to capture such an effect. In other regards, however, the belief model used is quite clearly artificial. Belief change is by simple averaging of information contacts, and all agents follow the same formula for belief updating. Our attempt in future work will be to test the robustness of conclusions here by considering a range of variations on the central model of belief change.

REFERENCES

- Auld, M. C. (2003). Choices, beliefs, and infectious disease dynamics. *Journal of Health Economics* 22, 361-377.
- Barbasi, A.-L. & R. Albert (1999). Emergence of scaling in random networks. *Science* 286, 509-512.
- Barrett, C. L., K. Bisset, J. Leidig, A. Marathe, & M. Marathe (2009). Estimating the impact of public and private strategies for controlling an epidemic: A multi-agent approach. *Proceedings of the Twenty-First Innovative Applications of Artificial Intelligence Conference, AAAI*, 34-39.

- Börner, K., Maru, J. T., & Goldstone, R. L. (2003). The simultaneous evolution of author and paper networks. *Proceedings of the National Academy of Sciences of the USA*, 101 (Suppl.1), 5266-5273.
- Centola, D., & M. Macy (2007). Complex contagion and the weakness of long ties. *American Journal of Sociology* 113, 702-34.
- Del Valle, S., H. Hethcote, J. M. Hyman, & C. Castillo-Chavez (2005). Effects of behavioral changes in a smallpox attack model. *Mathematical Biosciences*, 195, 228-251.
- Epstein, J. M., J. Parker, D. Cummings, & R. A. Hammond (2008). Coupled contagion dynamics of fear and disease: Mathematical and computational explorations. *PLoS ONE* 3(12): e3955. doi:10.1371/journal.pone.0003955
- Eubank, S., H. Guclu, V. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, & N. Wang (2004). Modeling disease outbreaks in realistic urban social networks. *Nature* 429:180-184.
- Fell, D., & A. Wagner (2000). The small world of metabolism. *Nature Biotechnology* 18, 1121-1122, reprinted in Newman, Barbási, & Watts, *The Structure and Dynamics of Networks*, Princeton: Princeton University Press, 2006.
- Faloutsos, M, P. Faloutsos, & C. Faloutsos (1999). On power-law relationships of the internet topology. SIGCOMM '99, reprinted in Newman, Barbási, & Watts, *The Structure and Dynamics of Networks*, Princeton: Princeton University Press, 2006.
- Ferrari, M. J., S. Bansal, L. A. Meyers & O. N. Bjørnstad (2006). Network frailty and the geometry of herd immunity. *Proceedings of the Royal Societ. B* 273, 2743-2748
- Freeman, L. C. (1978). Segregation in social networks. *Sociological Methods and Research* 6: 411-429.
- Funk, S., E. Gilad, C. Watkins, & V. A. A. Jansen (2009). The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences* 106 (16), 6872-6877, www.pnas.org / cgi / doi / 10.1073 / pnas.
- Golub, B., & M. O. Jackson, forthcoming. How homophily affects learning and diffusion in networks.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology* 78, 1360-1380.
- Grim, P., R. Rosenberg, A. Rosenfeld, B. Anderson, & R. Eason (2010). How simulations fail. Group for Logic & Formal Semantics research report #10-02, SUNY Stony Brook.
- Hallett, T. B., S. Gregson, J. J. C. Lewis, B. A. Lopman, & G. P. Garnett (2007). Africa: Behavioral change in generalized HIV epidemics: Impact of reducing cross-generational sex and delaying age at sexual debut. *Sexually Transmitted Infections* 83, p. i50-i54.
- Hamilton, D., M. Handcock & M. Morris (2008). Degree distributions in sexual networks: A framework for evaluating evidence. *Sexually Transmitted Diseases* 35, 30-40.
- Harrison, J. A., P. D. Mullen, & L. W. Green (1992). A meta-analysis of studies of the health belief model. *Health Education Research* 7:107-116.
- Janz, N. K., & M. H. Becker (1984). The health belief model: A decade later. *Health Education Quarterly* 11:1-47.
- Jeong, H., B. Tombor, R. Albert, Z. N. Ottvai, & A.-L. Barbási (2000). The large-scale organization of metabolic networks. *Nature* 407, 651-654, reprinted in Newman, Barbási, & Watts, *The Structure and Dynamics of Networks*, Princeton: Princeton University Press, 2006.
- Keeling, M., 2005. The implications of network structure for epidemic dynamics. *Theoretical Population Biology* 67, 1-8.
- Kincaid, D. L. (2000). Mass media, ideation, and behavior: A longitudinal analysis of contraceptive change in the Philippines. *Communication Research* 27, 723-763.
- Klov Dahl, A. S. (1985). Social networks and the spread of infectious diseases: The AIDS example. *Social Science & Medicine* 21, 1203-1216.

- Liljeros, F., C. R. Edling, L. A. Nunes Amaral, H. E. Stanley, & Y. Åberg (2001). The web of human sexual contacts. *Nature* 411, 907-908, reprinted in Newman, Barbási, & Watts, *The Structure and Dynamics of Networks*, Princeton: Princeton University Press, 2006.
- Meyers, A. M., B. Pourbohloul, M. E. J. Newman, D. M. Skowronski & R. C. Brunham 2005. Network theory and SARS: predicting outbreak diversity. *Journal of Theoretical Biology* 232, 71-81.
- Miller, J. C. & J. M. Hyman, Effective vaccination strategies for realistic social networks. *Physica A* 386, 780-785.
- Morris, M. (1997). Sexual networks and HIV. *Aids* 11 (suppl A), S209-S216.
- Morris, M., C. Podhisita, M. J. Wawer & M. S. Handcock (1996). Bridge populations in the spread of HIV/AIDS in Thailand. *Aids* 10, 1265-1271.
- Mullen, P. D., J. Hersey, & D. C. Iverson (1987). Health behavior models compared. *Social Science and Medicine* 24:973-981.
- Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98 (2), 404-409, reprinted in Newman, Barbási, & Watts, *The Structure and Dynamics of Networks*, Princeton: Princeton University Press, 2006.
- Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics* 46: 323-351.
- Redner, S. (1998). How popular is your paper? An empirical study of the citation distribution. *European Physical Journal B* 4, 131-134.
- Rosenberg, R., P. Grim, A. Rosenfeld, B. Anderson & R. Eason (2010). The Science in simulation: a structural analysis. Group for Logic & Formal Semantics research report #10-01, SUNY Stony Brook.
- Strecher, V. J., & I. M. Rosenstock (1997). The health belief model. In *Health Behavior and Health Education: Theory, Research, and Practice*, eds. K. Glanz, F. M. Lewis, and B. K. Rimer. San Francisco: Jossey-Bass.
- Trotter, R. T., R. B. Rothenberg, & S. Coyle (1995). Drug abuse and HIV prevention research: Expanding paradigms and network contributions to risk reduction. *Connections* 18, 29-45.
- Valente, T. (1995). *Network Models of the Diffusion of Innovations*. Cresskill, N. J.: Hampton Press.
- Valente, T. (2010). *Social Networks and Health: Models, Methods, and Applications*. New York: Oxford University Press.
- Valente, T., & R. L. Davis (1999). Accelerating the diffusion of innovations using opinion leaders. *Annals of the AAPSS* 566, 55-67.
- Watts, D. J., & S. H. Strogatz (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440.

