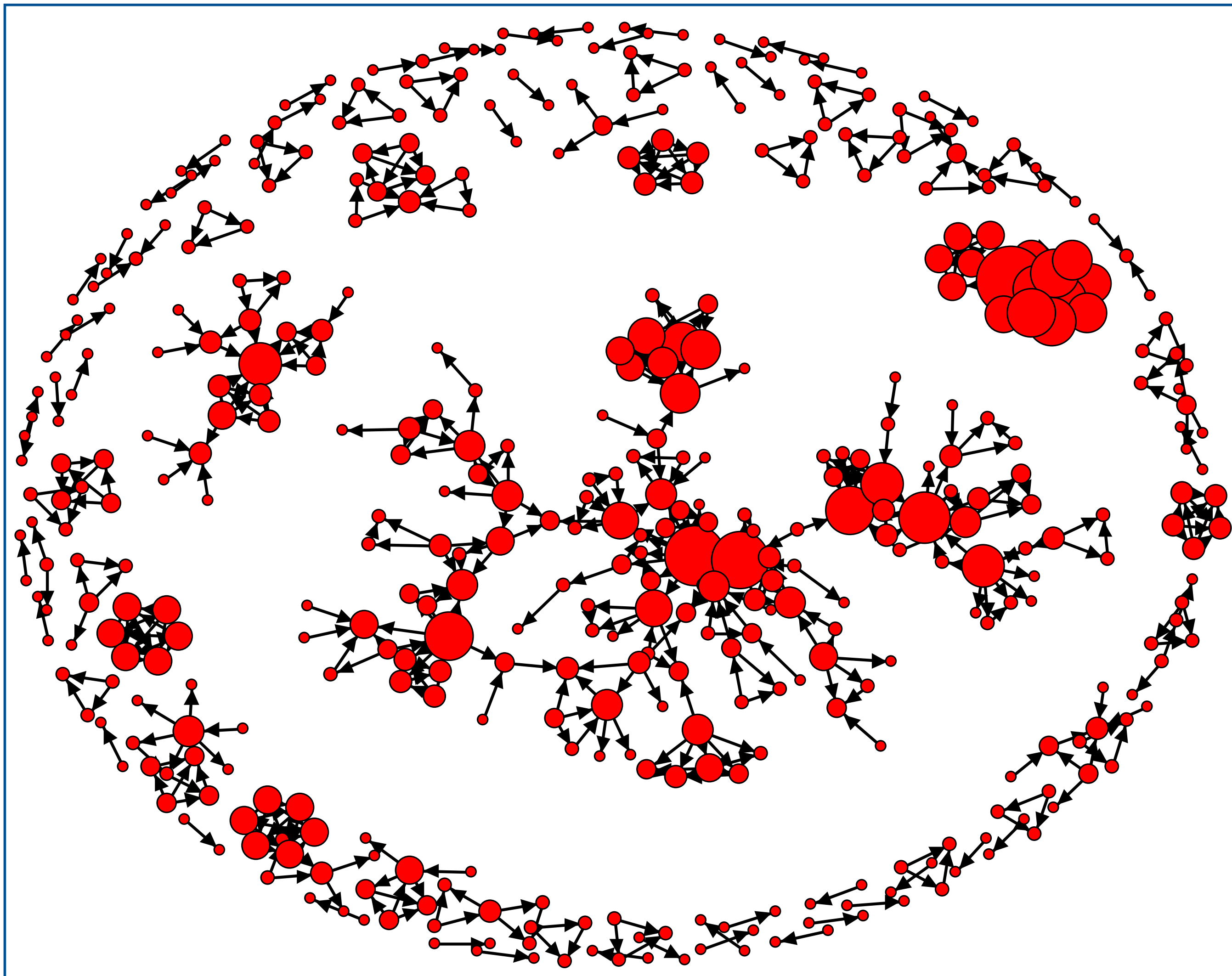
CONNECTIONS

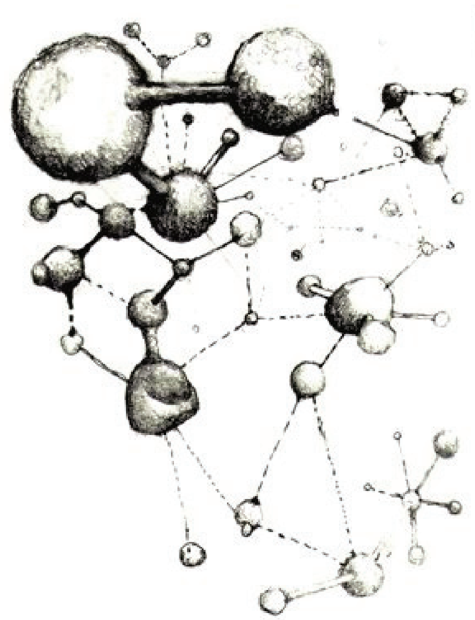
June 2008 *Volume 28 • Issue 1*



Inside this issue:

Official Journal of the International Network for Social Network Analysis

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, Organizational Behavior, Knowledge Management, Marketing, Social Psychology, Public Health, Medicine, Computer Science, 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 news about social network concepts and techniques and new tools for research.



***Front Cover*:** Created by Chris McCarty, the graph represents all authors who have ever published in the journal Social Networks, excluding those who only published as single authors (isolates). Data were downloaded from the ISI Web of Science. Nodes represent authors, node size represents degree centrality, and the thickness of the line between nodes represents the number of articles co-authored in Social Networks. The design was used for the last Sunbelt conference (Sunbelt XXVIII International Sunbelt Social Network Conference) in January 2008.

<u>International Network for Social Network Analysis</u>

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. INSNA includes official board members and five committees.

Board Members

George A. Barnett, President
Pip Pattison, Vice President
Thomas W. Valente, Treasurer
Barry Wellman, Founder
Phil Bonacich, Martin Everett, Katie Faust, Scott Feld, Anuska Ferligoj, Garry Robins, John Skvoretz, Stan Wasserman

Committees

<u>Finance Committee</u>**,** Chaired by Treasurer Thomas W. Valente
<u>Program/Conference Committee</u>**,** Co-chaired by the host(s) Chris McCarty and John Skvoretz of the Sunbelt Conference and Russ Bernard
<u>Web Committee</u>, Chaired by a webmaster (chief information officer) Benjamin Elbirt.
<u>Publications Committee</u>, Chair, TBD. Composed of current and former and 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 complementary rather than in competition with one another. To insure openness to new ideas, one or additional members will be selected by the President.
<u>Awards Committee</u>**,** The awards committee would be composed of four sub-committees:

- Visual Path Award, Dan Brass
- Freeman Award, Noshir Contractor
- Simmel Award, Russ Bernard, Chris McCarty, & John Skvoretz
- Microsoft Award**,** Janet Fulk, & Mario Diani

CONNECTIONS
Manuscripts selected for publication are done so based on a peer-review process. See 'Instructions for Authors' for information on submitting manuscripts for publication. The journal is edited and published by Editor, Thomas W. Valente (Director of the Master of Public Health Program and Professor in the Department of Preventive Medicine at the University of Southern California) and Managing Editor, Kathryn Coronges, (PhD Student of Health Behavior Research at the University of Southern California). Editorial headquarters are located at USC's Institute of Prevention Research, 1000 Fremont Ave., Unit #8, Building A, Room 5133, Alhambra, CA 91803. Tel: (626) 457-6678; fax: (626) 457-6699. Email tvalente@usc.edu or coronges@usc.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)

Wherever possible, items referenced in articles (such as data and software) are made available electronically via INSNA's website, http://www.insna.org. The website provides access to a directory of members' email addresses, network datasets, software programs, and other items that lend themselves to electronic storage. In addition, the website provides updated information on the annual **International Sunbelt Social Network Conferences** where researchers come together to share current theoretical, empirical and methodological outlooks and findings.

Hardcopy circulation of Connections is sent to all INSNA members, which has reached almost twelve hundred. Subscription to CONNECTIONS can be obtained by registering for INSNA membership through the website. Standard membership fee is US$50 ($35 for students, $60 for institutions).

CONNECTIONS

June 2008 **Volume 28 • Issue 1**

CONTENTS

ARTICLES

Instructions to Authors

CONNECTIONS publishes original empirical, theoretical, tutorial, and methodological articles that use social network analysis. The journal seeks to publish significant work from any domain that is relevant to social network applications and methods. Review articles that critically review and synthesize a body of published research, are normally by invitation only. Commentaries or short papers in response to previous articles published in the journal are also considered for publication. Authors who wish to submit a commentary, book review, network image or review article should first e-mail the editors with a brief description. Overlap between submitted manuscripts and published abstracts will be allowed, provided that such abstracts are not verbatim reproductions of the previously published abstract.

Submitting Manuscripts

Authors are required to submit manuscripts online to the editor, Thomas W. 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 or WordPerfect format and should not exceed 40 pages including tables, figures and references. Manuscripts should be arranged in the following order: title page, acknowledgments, abstract, text, references, appendices, and figure legends. 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*. Include author's contact information in the title page. Abstracts should be limited to 250 words. Please embed all images, tables and figures in the document. If you have a large figure, you may also send it as a separate file. A figure and its legend should be sufficiently informative that the results can be understood without reference to the text.

Rethinking Preferential Attachment Scheme: Degree centrality versus closeness centrality

Kilkon Ko[1]
University of Pittsburgh, USA

Kyoung Jun Lee
Kyung Hee University, Korea

Chisung Park
University of Pittsburgh, USA

Abstract:

Construction of realistic dynamic complex networks has become increasingly important. One of the more widely known approaches, Barabasi and Albert's "scale-free" network (BA network), has been generated under the assumption that new actors make ties with high degree actors. Unfortunately, degree, as a preferential attachment scheme, is limited to a local property of network structure, which social network theory has pointed out for a long time. In order to complement this shortcoming of degree preferential attachment, this paper not only introduces closeness preferential attachment, but also compares the relationships between the degree and closeness centrality in three different types of networks: random network, degree preferential attachment network, and closeness preferential attachment network. We show that a high degree is not a necessary condition for an actor to have high closeness. Degree preferential attachment network and sparse random network have relatively small correlation between degree and closeness centrality. Also, the simulation of closeness preferential attachment network suggests that individuals' efforts to increase their own closeness will lead to inefficiency in the whole network.

---

[1] Direct correspondent to Kilkon Ko, 3601 Posvar Hall, Graduate School of Public and International Affairs, University of Pittsburgh, PA 15260, USA, Email:kilkon@gmail.com

We are grateful to anonymous reviewers who help us to clarify concepts and assumptions. This research is supported by the Ubiquitous Autonomic Computing and Network Project, the Ministry of Information and Communication (MIC) 21st Century Frontier R&D Program in Korea

INTRODUCTION

During the last decade, a considerable number of empirical studies have suggested that "scale-free" is one of the most conspicuous structural property in large complex networks (Barabasi, 2002; Buchanan, 2002; Newman, 2003; Strogatz, 2004; Watts, 2003). A scale-free network is a network whose degree distribution follows a power law, i.e. that the probability of having a node with degree k satisfies $P(k)=k^{-\gamma}$. Common characteristics of the scale-free network are 1) centrally located and interconnected high degree "hubs," 2) small average distance among nodes, and 3) high clustering coefficient (Barabasi, 2002; Watts, 1999).

While many mechanisms can be used for simulating the scale-free structure (Keller, 2005), the degree preferential attachment assumption has been widely used in both mathematical (Dorogovtsev & Mendes, 2002) and simulation approaches (Albert *et al.*, 2000; Barabasi & Albert, 1999; Barabasi, 2002; Dorogovtsev & Mendes, 2002). The attachment rule assumes that actors try to make a tie with other actors who maintain high degree centrality.

But for the popularity of preferential attachment, yet few have questioned whether degree-based preferential attachment is appropriate. Social network theory, however, has recognized that indirect relations can be more important than the direct ones. As the degree centrality focuses on the local property of the network structure and overemphasizes the direct relations (Freeman, 1979), it underestimates the importance of indirect relations or the global property of the network structure. For this reason, social network theory has developed additional measures such as betweenness, closeness, information or power centrality. Literatures in scale-free network, however, have not dealt with the appropriateness of degree preferential attachment. The assumption might not be a problem if degree centrality is highly correlated to other centrality measures. To our knowledge, however, there is no study that suggests positive or negative relationships between degree centrality and other centrality measures.

This study aims to bridge this knowledge gap. When degree centrality was introduced by Shaw (1954), it was so intuitively appealing that network researchers admitted this concept as a fundamental feature to explain the network structure. If an actor occupies a structural location to connect with other actors with many adjacent ties, she can be seen as a major channel of information in communication network. For example, in the friendship network, if a person receives many choices from others, she can be considered as a focal point of friendship network.

However, as Granovetter (1973) raises the importance of the indirect ties (i.e., the strength of weak ties), social network theory becomes interested in the fact that a high degree is not a necessary condition to be a powerful actor. Unlike Winship's hierarchy model for group classification (1977), Granovetter's weak tie model looses the condition of intransitivity (Freeman, 1992). It argues that if A and B have strong ties and B and C have strong ties, then it will be enough to assume that A and C are weakly linked. The structural location of a weak tie can play an important role in this intransitivity. Burt (2000) argues that if a social network is composed of multiple groups which are internally cohesive, many actors can have a high degree and short distance within their cohesive groups.

However, an actor who bridges two internally cohesive groups - but externally weakly linked between the two cohesive groups – would have more opportunities in getting information as well as mobilizing embedded resources in a timely manner.

Such insights of social network theory make us rethink the attachment rule. If actors of low degree can reach each other with a small number of paths, the comparative advantage of high degree actors in its accessibility to others will not be large. Actors may prefer to choose an influential actor(s) in order to minimize their costs to access information or embedded resources. Thus, the "influential actors" can be actors who have high degree and/or high closeness centrality. When degree and closeness centrality are not highly correlated, we can not say the one is always better than the other (Freeman, 1979). But most studies on current scale-free networks (Barabasi, 2002; Newman, 2001; Watts, 2003) have used degree centrality as an attachment scheme without providing detailed grounds. In other words, the existing studies do not pay much attention to the relationship between the degree and closeness centrality in modeling dynamic complex network. The main goal of this study is not to discuss the superiority between degree and closeness as a centrality measure. Instead, this paper attempts to answer the following questions:

i) Are degree and closeness centrality highly correlated in scale-free networks?

ii) If they are (not) correlated, under what conditions are they (not) correlated?

iii) If we use closeness centrality as an attachment scheme in modeling scale-free networks, how will results be distinguished from the scale-free networks using degree centrality?

In order to answer the above questions, this study simulates[2] the two different types of networks commonly used to define the network topology: random network and degree preferential attachment network. We compare the two networks by looking at the relationship between degree and closeness. In addition to these two networks, we also introduce *a closeness preferential attachment network,* which shows the structural difference to the other two networks.

To preserve the comparability among different networks, we simulate those three different types of networks having the same number of actors (N=500) and ties (≈1000). According to random graph theory and empirical studies on large networks, the low density networks are more likely to be observed. In order to achieve reliable results, we repeated the simulation 50 times for each type of networks. Since the goal of this paper is to explore new knowledge that is understudied in the literature, we do not attempt to perform a full-scale simulation.

In the following sections, we examine three different ways of generating dynamic networks: *random network, degree preferential attachment network* (hereafter DPN), and *closeness preferential attachment network* (hereafter CPN). After providing simulation results on the relationship between degree and distances, and the structural difference among them, we will present the implications and conclusion.

[2] We used SAS 8.2 for simulation and data analyses.

Distance and Three Types of Network

Within a network, several paths may exist between a pair of nodes. In that case, the shortest path between two nodes is called the geodesic distance (Wasserman and Faust, 1994). This study uses the geodesic distances as a distance measure for the closeness centrality (Valente & Foreman, 1998). A geodesic distance is infinite if two nodes are not connected to each other. Although one simple way is to consider only pairs of reachable nodes, the weakness of this approach is to underestimate the role of isolated nodes. A better alternative to define the distance between unreachable nodes is to use a size of network, which is used in UCINET and also in this paper. If the network size is *N*, then the maximum path to reach other node will be *N*-1.

The closeness of an actor is measured with the actor's closeness centrality. Let's denote a geodesic distance between two nodes n_i, n_j as $d(n_i, n_j)$. The actor's closeness centrality is the inverse of the sum of geodesic distances from actor *i* to the *N*-1 other actors.

$C_c(n_i) = [\sum_{j=1}^{N} d(n_i, n_j)]^{-1}$, where *i≠j*

Also the normalized closeness centrality controlling the network size effect is defined as:

$C^N{}_c(n_i) = (N-1) * [\sum_{j=1}^{N} d(n_i, n_j)]^{-1}$, where *i≠j*

The average of the normalized closeness centrality of all nodes is the reciprocal of the average distance of the network.

Random network

Random networks are the simplest, but most widely discussed network form (Chung & Lu, 2001; Erdos & Renyi, 1960). Just as many other real world processes have been effectively modeled by appropriate random models, a random network provides useful insights to understand complex networks (Aiello *et al.*, 2000). We can generate a random network in two different ways. One way is to start with couple of nodes at an initial stage and have new nodes connected to existing nodes randomly. The other way is to create a network connecting two arbitrary nodes with equal probability, *p*, after fixing the total number of nodes (*N*). While the former approach is more appropriate for describing dynamic growth of a network, the latter classical random network is easier to mathematically operationalize. Thus, we use the classical random network in the following:

Theoretically, the random network will have a total number of links: $L = p * \frac{N(N-1)}{2}$.

The degree distribution, a probability that a certain node has *k* degree, follows binomial distribution defined: $P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$. And the average degree is $\bar{k} = p(N-1) \approx pN$. If the N is large enough, the degree distribution will follow the Poisson distribution, $P(k) = e^{-\bar{k}} \bar{k}^k / k!$, which approximates a normal distribution when N is large. The density of random network is the same to *p* because density is defined as $\Delta = \frac{L}{N(N-1)/2} = \frac{2L}{N(N-1)}$.

The average distance of random network will be $\bar{\lambda} \sim \ln N / \ln[pN]$ (Dorogovtsev & Mendes, 2002:15). Compared to degree and density, the distance of random network is not affected so much by the size of network (*N*). To examine the impact of the size of network and the probability of attachment, we simulate the random network 50 times by changing N and p. After controlling the p, we analyze the impact of network size on closeness centrality measured as the reciprocal of distances. Figure 1 shows that the impact of the size of the network closeness decreases significantly as the size of the network grows.

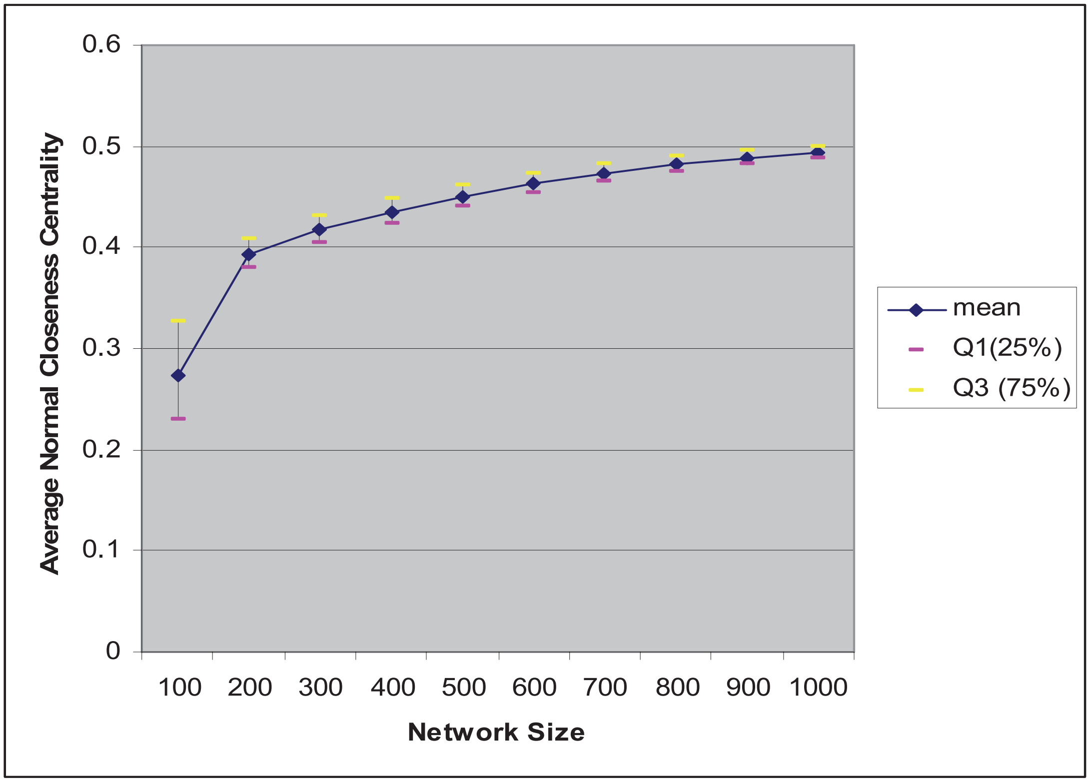


Figure 1. The positive relationship between size of network and the normalized closeness.

In contrast, average distance is a decreasing function of p. As the first order derivative of average distance with respect to p is $\frac{Log(N)}{pLog(NP)^2}$, average distance is a monotonic decreasing function of p. After controlling the size of network, we simulate a random network to observe the impact of the p on closeness centrality. As the probability of a link increases, the average distances between nodes decrease as shown in Figure 2.

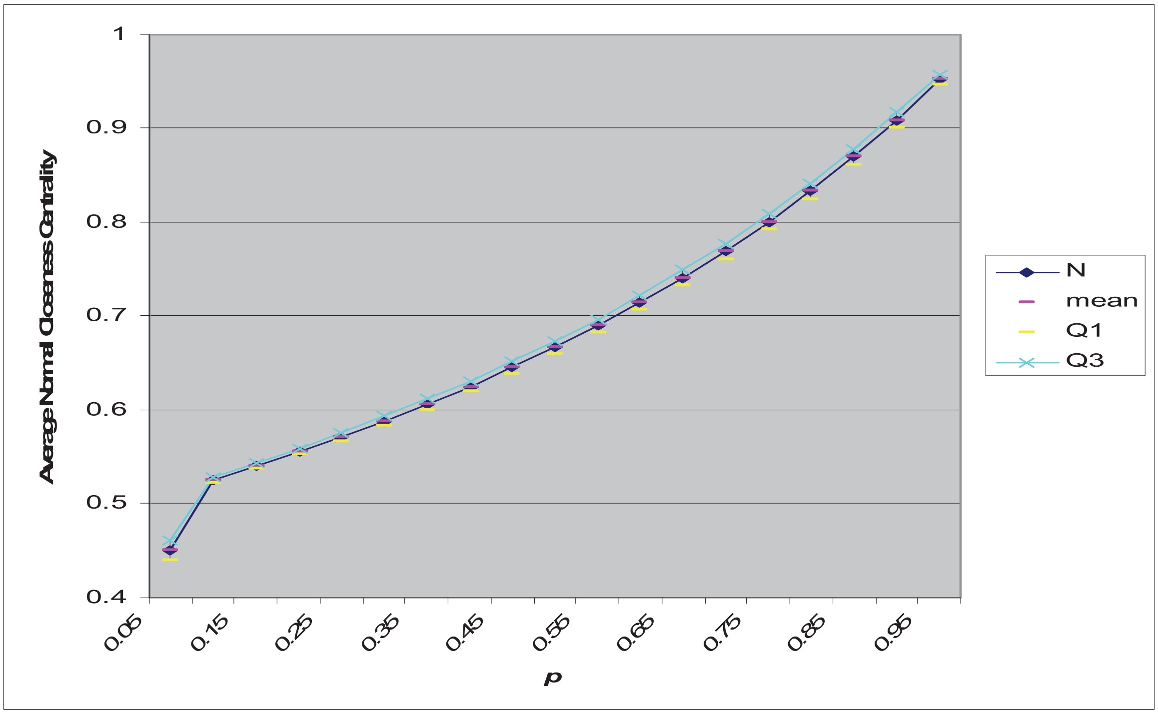


Figure 2. The impact of probability of attachment on normalized closeness.

Finally, we analyze the relationship between degree and closeness. The relationship between degree and closeness can be measured by the Pearson correlation coefficient. Figure 3, shown below, illustrates the correlation between closeness centrality and degree centrality of actor by changing *p* from 0.005 to 0.101. The correlation is high when *p* exceeds a critical value (greater than 0.025 in our simulation) regardless of the size of the network. In contrast, when *p* is small, the correlation coefficient is small regardless of the size of the network.

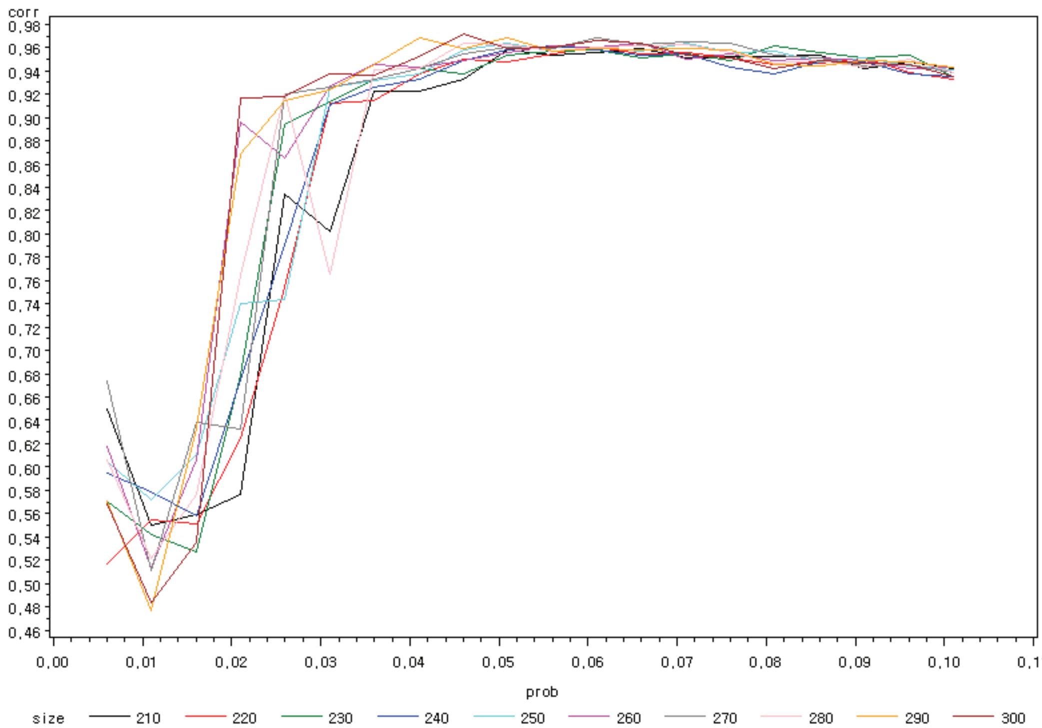


Figure 3. Correlation between closeness and degree centrality in random network.

An interesting question at this point is whether a network which has small correlation coefficient between degree and closeness centrality is rare in the real network or not. The question is related to the size of probability of making ties between actors (i.e. *p*). According to empirical studies (Newman, 2003), the average distance of a social network such as film actors (N=449,913), CEOs, academic co-authorships (N=52,905~253,339), and e-mail message networks (N=59,912) ranges from 3.48 to 16.01. We can infer the probability of attachments, *p*, using the size of network and the average distance. As the average distance of random network is $\bar{\lambda} \sim \ln N / \ln[pN]$, the large *p* (e.g. p=0.101) will lead less than a 2 degree average distance in the network when *N* is greater than 100. Even if we have the small average distance such as 3.48 from the network size N=100,000, *p* is 0.0003, which is small enough to have small correlation coefficients.[3] Thus, networks whose degree and closeness are weakly correlated can be frequently found even though the small world effect exists. In sum, degree and closeness centrality are not correlated in low density networks.

[3] Of course, the empirical networks are not random networks. So, the probability of making ties based on random network assumption might not be applicable. However, when we measured the probability of making ties in empirical networks using the size of the network and the number of edges, we also get very small probability. For instance, a film actor network whose average distance is 3.48, has 449,913 nodes and 25,516,482 edges. As the probability of making ties are the same to density, p is 0.000126.

Degree Preferential Attachment Network (DPN)

During the last decade, a preferential attachment has been regarded as a basic rule governing the formation and evolution of real networks (Dorogovtsev & Mendes, 2002; Newman, 2003; Watts, 2003). The widely used algorithms for generating the preferential attachment network are (Barabasi, 2002:86):

i) Growth: For each given period of time, adding a new node to the network.
ii) Degree preferential attachment: Each new node connects to the existing nodes with two ties. The probability that it will choose a given node is proportional to a degree the chosen node has.

This paper starts from a highly cohesive seed network with size 10, having the probability of making ties, 0.9, and following the same rule of degree preferential attachment. [4] The most distinguishing feature as compared to the random network is that the degree preferential attachment network has a heavier tail than the random network. It has been known that the probability that DPN has a node with k degrees follows a power law with a form of $p(k) \sim k^{-\alpha}$. Although the fluctuations in the tail of degree distribution are not small (Dorogovtsev & Samukhin, 2002), our simulation produces pretty stable maximum degree distribution. DPN and the random network have a similar average degree, 2, because, for a random network, the expected average degree is Np=500*0.004=2. Also our simulation of DPN has an average degree 2.1. However there exists an actor having maximum 45 in-degrees in DPN, whereas there is an actor having less than 5 in-degrees in the random network.

To compare the efficiency of network, we calculate the average distance. As the long average distance implies the less efficient in accessibility, we use the term 'efficiency' to refer to the closeness in this paper. The average distance of the random network is 8.97 but DPN has 4.6. If the size of the network and the number of ties are equal, we can confirm DPN is more efficient than the random network.

DPN also does not have a strong correlation between degree and closeness centrality. If high degree does not guarantee high closeness, it brings out an important question: what if actors want to make a tie to others who have higher closeness rather than a high degree? We will call this preferential attachment scheme as a closeness preferential attachment.

Closeness Preferential Attachment Network (CPN)

Unlike DPN, CPN uses closeness as a weight of attachment instead of degree. It follows, then, a new actor makes ties with high closeness actors with high probability. Other conditions are the same to those used in DPN.

Under the assumption of high correlation between degree and closeness centrality, the network structure of DPN and CPN should not be different. DPN and CPN, however, do show different structures. First, the correlation between degree and closeness in CPN is larger than that in DPN. As seen in

[4] This paper repeats the simulation 50 times to control the random effects.

Table 1, the correlation coefficient between in-closeness and in-degree is 0.47 in DPN, but 0.67 in CPN. This result may be because the attachment rule of CPN is based on the closeness centrality.

Table 1. Correlation coefficients between closeness and degree

		In Degree	Out Degree
In Closeness	DPN	0.47325	0.43704
	CPN	0.67421	0.56243
Out Closeness	DPN	0.42947	0.46331
	CPN	0.55853	0.66811

Note: All coefficients are statistically significant under alpha=0.05.

Second, the average distance of CPN is larger than that of DPN as shown in Table 2 below. As DPN has the same number of actors and ties, the longer distance of CPN implies that CPN is less efficient than DPN.

Table 2. Average distance of Networks

	In distance	Out distance
Random Network[5]	8.97	
DPN	4.46 (0.86)	4.46 (0.97)
CPN	6.43 (1.03)	6.43 (1.22)

Note: Standard deviation is in the parenthesis, the size of network is 500.

[5] As we construct non-directional random network, the in and out degree are the same. In case of directional random graph, our finding is consistent.

The result is paradoxical in that we assume that individual actors try to make ties based on closeness centrality to reduce distance to others. While individuals make their ties to high closeness actors to reduce his or her distance to others, these efforts of each individual, however, are not transformed into increasing efficiency of the overall system.

Such result comes from the fact that the closeness centrality is sensitive to other actors' structural position. The closeness centrality is interdependent to other actors and ties. It does not simply depend on one actor's own choice. When the dynamic network is constructed, network is not evenly distributed. Compared to random networks, DPN and CPN have higher clustering coefficients which imply the emergence of subgroups within a network. When internally cohesive subgroups emerge, an actor with small degree, but high closeness centrality, such as a cutpoint, becomes important in the network. Although DPN puts little weight on the cutpoint in the attachment process, CPN puts much weight on it. The structural advantage of a cutpoint in CPN, however, will decrease as new ties are added between two subgroups. As a result, CPN will have more equally distributed degree distribution than DPN. The maximum degree of DPN and CPN in Table 3 shows that CPN can not have an extremely high degree node.

Table 3. Maximum In and Out Degree of Networks

	In Degree	Out Degree
Random Network	5	
DPN	45	52
CPN	18	20

In sum, compared to DPN, CPN has higher correlation coefficient of degree and closeness centrality. Individual actors make their own best choices by connecting themselves to high closeness centrality actors and reduce distance to others. However, the individual's best choice does not guarantee the whole network's efficiency. As our results show, CPN has lower efficiency than DPN. The low efficiency of CPN mainly comes from the fact that the high degree actors in CPN have relatively small numbers compared to their neighbors.

DISCUSSION

While BA's "scale-free" network has used degree centrality as a basic preferential attachment rule, it is only one of possible measures of important actors (Bonacich, 1987; Scott, 2000; Stephenson & Zelen, 1989; Wasserman & Faust, 1994). Our study shows that high degree nodes are not necessarily high closeness centrality nodes in a sparse random network or a degree preferential attachment network. In particular, the correlation between degree and closeness centrality in degree preferential attachment network is not so large.

We also simulated and compared the structural features of three different types of networks: random network, DPN (degree preferential attachment network), and CPN (closeness preferential attachment network). The results revealed that CPN has distinctive characteristics, different from both DPN and random network. CPN showed relatively higher correlation between degree and closeness centrality than DPN and random networks. A noticeable finding is that CPN had longer average distances than DPN, although individual actors in CPN tried to minimize distance to others. Within a closeness-oriented scheme, highly centered actors are less likely to emerge than in a degree-oriented network, as shown in Table 4.

Table 4. Summary of three models by average distance and maximum degree

		Average distance		
		Low	Medium	High
Maximum degree	Low	–	–	Random Network
	Medium	–	CPN	–
	High	DPN	–	–

Although scale-free networks are a useful framework to understand the property of complex networks such as hyperlink network, power grid, telecommunication, internet, or biological network, it is not clear how such networks are formed and evolve. The ambiguity of evolutionary mechanism is mainly due to the lack of knowledge about the possible impetus of networking. As our study proposes, one may prefer to be connected to high degree nodes, but others may prefer to be connected to high closeness nodes. Thus, in future studies, we should pay more attention to the relationship among various measures of preferential attachment. At the same time, we have to analyze how different attachment schemes generate different network structures.

References

Aiello, W., Chung, F., & Lu, L. (2000). *A random graph model for massive graphs.* Paper presented at the The 32d Annual ACM Symposium on Theory of Computing.

Albert, R., Jeong, H., & Barabasi, A.-L. (2000). Error and attack tolerance in complex networks. *Nature, 406*, 378-381.

Barabasi, A.-L. (2002). *Linked:The new science of networks*. Cambridge:MA: Perseus Publishing.

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

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology, 92*, 1170-1182.

Buchanan, M. (2002). *Nexus: Small worlds and the groundbreaking science of networks* (1st ed.). New York: W.W. Norton.

Burt, R. S. (2000). The network structure of social capital. In R. I. Sutton & B. M. Staw (Eds.), *Research in organizational behavior, volume 22*. Greenwich, CT: JAI Press.

Chung, F., & Lu, L. (2001). The diameter of random sparse graphs. *Advances in Applied Math., 26*, 257-279.

Dorogovtsev, S. N., & Mendes, J. F. F. (2002). Evolution of networks. *Advances In Physics, 51*, 1079-1146.

Dorogovtsev, S. N., & Samukhin, A. N. (2002). Mesoscopics and fluctuations in networks. *cond-mat/0211518*.

Erdos, P., & Renyi, A. (1960). On the evolution of random graphs. *Publ. Math. Inst. Internat. Acad. Sci., 5*, 17-61.

Freeman, L. C. (1979). Centrality in social networks conceptual clarification. *Social Networks, 1*(3), 215-239.

Freeman, L. C. (1992). The sociological concept of "group": An empirical test of two models. *The American Journal of Sociology, 98*(1), 152-166.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360-1380.

Keller, E. F. (2005). Revisiting scale-free networks. *BioEssays, 27*(10), 1060 - 1068.

Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E, 64*(025102).

Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*, 167-256.

Scott, J. (2000). *Social network analysis: A handbook* (2nd ed.). London; Thousands Oaks, Calif.: SAGE Publications.

Shaw, ME. (1954). Some effects of unequal distribution of information upon group performance in various communication nets. *J Abnorm Psychol*. 49(1, Part 1): 547-53.

Stephenson, K., & Zelen, M. (1989). Rethinking centrality: Methods and examples. *Social Networks, 11*(1), 1-37.

Strogatz, S. H. (2004). *Sync: The emerging science of spontaneous order*. Hyperion.

Valente, T. W., & Foreman, R. K. (1998). Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Social Networks, 20*, 89-105.

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

Watts, D. J. (1999). *Small worlds: The dynamics of networks between order and randomness*. Princeton, N.J.: Princeton University Press.

Watts, D. J. (2003). *Six degrees: The science of a connected age* (1st ed.). New York: W.W. Norton.

Winship, C. (1977). A Distance Model for Sociometric Structure. *Journal of Mathematical Sociology* 5: 21-39.

How Correlated Are Network Centrality Measures?

Thomas W. Valente, PhD
University of Southern California, Department of Prevention Research, Los Angeles

Kathryn Coronges, MPH
University of Southern California, Department of Prevention Research, Los Angeles

Cynthia Lakon, PhD
University of Southern California, Department of Prevention Research, Los Angeles

Elizabeth Costenbader, PhD
Research Triangle Institute, Raleigh North Carolina

Corresponding author: Prof. Thomas Valente, Department of Preventive Medicine, School of Medicine, University of Southern California, 1000 Fremont Ave, Bldg A Room 5133, Alhambra CA 91803 ;email : tvalente@usc.edu)

We thank Phil Bonacich who inspired this paper and provided helpful comments on earlier drafts, and Noah Friedkin for SNAPS which was used to calculate the centrality measures. Support for this research was provided by NIDA grant P50-DA16094.

INTRODUCTION

Calculating centrality has been a major focus of social network analysis research for some time (Freeman, 1979). Textbooks and reference volumes on social networks include a chapter on centrality calculations and concepts (e.g., Degenne & Forsé, 1999; Scott, 2000; Wasserman & Faust, 1994). Currently, at least eight centrality measures have been proposed and made available in UCINET 6 (Borgatti, et al., 2005). These measures are: degree, betweenness, closeness, eigenvector, power, information, flow, and reach.

Perhaps the most frequently used centrality measures are degree, closeness, betweenness, and eigenvector. The first three were proposed by Freeman (1979) and eigenvector was proposed by Bonacich (1972). Centrality is important because it indicates who occupies critical positions in the network. Central positions have often been equated with opinion leadership or popularity, both of which have been shown to be associated with adoption behaviors (Becker, 1970; Rogers, 2003; Valente, 1995; Valente & Davis, 1999). Typically, investigators use only the degree measure of centrality (simply the number of links a person has), as it is the easiest to explain to non-network savvy audiences and its association with behavior is intuitive.

An often asked, yet rarely answered question has been: Are these centrality measures correlated? All centrality measures are derived from the adjacency matrix and so constitute different mathematical computations on the same underlying data. If the measures are highly correlated, then the development of multiple measures may be somewhat redundant and we can expect the different measures to behave similarly in statistical analyses. On the other hand, if the measures are not highly correlated, they indicate distinctive measures likely to be associated with different outcomes.

Previous studies have examined correlations among centrality measures. One study examined correlations between degree, closeness, betweenness, and flow, and also examined these relationships under conditions of random error, systematic error, and incomplete data (Bolland, 1988). Overall degree, closeness, and continuing flow centrality were strongly intercorrelated, while betweenness remained relatively uncorrelated with the other three measures (Bolland, 1988). In a network study of individuals connected through participation in HIV risk behaviors, Rothenberg and colleagues (1995) examined relationships among eight centrality measures: three forms of information centrality, three distance measures (i.e., eccentricity, mean, and median), and degree and betweenness centrality. Their analyses showed these eight centrality measures to be highly correlated with a few notable distinctions. While the three distance measures were highly interrelated, they were also strongly correlated with the three information measures, although less so with degree and betweenness. The latter two measures, degree and betweenness, were highly correlated, although less so with information measures. The information measures were also highly correlated.

In another study, Valente and Forman (1998) examined correlations between measures of integration and radiality and other centrality measures and personal network density. Using data from the Sampson Monastery dataset (1969) and the Medical Innovations study (Coleman et al. 1966; Burt 1987), they found that integration was most highly and positively correlated with in-degree centrality, positively correlated with closeness, betweenness, and flow, and negatively correlated with density (Valente & Foreman, 1998). In comparison, radiality was significantly and negatively correlated with out-degree but only in the Medical Innovations dataset. Lastly, Faust (1997) examined correlations among centrality measures using a subset of the data from Galaskiewicz's study (1985) regarding relationships between CEOs, clubs and boards. Faust (1997) found correlations ranging from .89 to .99 among centrality measures including degree, closeness, betweenness, the centrality of an event, and flow betweenness for the identification of central clubs.

In this manuscript, we empirically investigate the correlation among four centrality measures, which we felt were those most commonly used by network analysts: degree, betweenness, closeness, and eigenvector. Degree and closeness are directional measures, so we calculate both in-degree and out-degree, and in-closeness and out-closeness. Closeness was calculated by inverting the distance matrix and taking the row average for closeness-out and the column average for closeness-in (Freeman, 1979). Nodes that were disconnected were given a distance of N-1 so that distances could be calculated. We also calculated closeness based on reversed distances (so called integration/radiality) but found these measures to be largely redundant with closeness based on inverting distances (Valente & Foreman, 1998). Betweenness indicates how frequently a node lies along geodesic pathways of other nodes in the network, and therefore is an inherently asymmetric measure. Eigenvector can only be calculated on a symmetric network and so matrices have to be symmetrized before eigenvector centrality is calculated. To compare eigenvector centrality to the other three measures thus requires that degree, closeness, and betweenness be calculated on symmetric data as well.

Degree, betweenness, eigenvector and closeness are all measure of an actor's prominence in a network (Wasserman & Faust, 1994). While considerable conceptual overlap exists between these constructs, they also may be conceptually distinct. For example, a node in the center of a star or wheel is the most central node in the network by all centrality measures (Freeman, 1979). In other network configurations, however, nodes with high degree centrality are not necessarily the most strategically located. One way to characterize such distinctions among these constructs is in terms of how actors who occupy positions high on each type of centrality transmit influence to other actors in a network.

We might expect that the pathway of influence transmitted from nodes high in degree and closeness centrality will be similar. Both can quickly transmit information and influence through direct or short paths to others and interact with many others directly. Closeness measures are based on the ideas of efficiency and independence (Freidkin, 1991). As a result of being situated close to others in the network, actors high on closeness measures are able to efficiently transmit information and have independence in the sense that they do not need to seek information from other more peripheral actors.

Betweenness centrality measures the extent to which an actor lies between other actors on their geodesics. Actors high on betweenness centrality, therefore, have the potential to influence others near them in a network (Friedkin, 1991), seemingly through both direct and indirect pathways. A node with high betweenness centrality can potentially influence the spread of information through the network, by facilitating, hindering, or even altering the communication between others (Freeman, 1979; Newman, 2003). Similarly, those high on eigenvector centrality are linked to well-connected actors and so may influence many others in the network either directly or indirectly through their connections.

We expect that measures of degree and closeness centrality will be more highly correlated with each other than with other measures, because they are both based on direct ties. We are unsure, however, how the other centrality measures will correlate with one another. Conceptually, each centrality measure represents a different process by which key players might influence the flow of information through a social network. In this study we examine the correlation between the symmetrized and directed versions of four centrality measures; symmetrized degree, in-degree, and out-degree, symmetrized betweenness, and betweenness, symmetrized closeness, closeness-in, and closeness-out, and eigenvector (symmetric only). We calculated these nine centrality measures for 58 existing social networks (from seven separate studies) analyzed previously by Costenbader and Valente (2003).

We correlated the 9 measures for each network and then calculated the average correlation, standard deviation, and range across centrality measures. We also calculated the overall correlation and compared it by study to assess the degree of variation in average correlation between studies. Lastly, we explore the associations between four different sociometric network properties (i.e., density, reciprocity, centralization and number of components) and the centrality correlations. This last analysis seeks to determine whether centrality measures are more highly correlated in dense or sparse networks, in reciprocal or non-reciprocal networks, in centralized or decentralized networks, and in networks with few or many components. Density is the number of ties in the network divided by the total possible number of ties (N*(N-1)). Reciprocity was measured as the percent of possible ties that are symmetric. Degree centralization was measured using Freeman's (1979) formula. The number of components in the network was determined by symmetrizing the network and calculating components.

METHODS

Data were originally collected in 7 studies, which included 62 sociometric networks in a variety of settings. All of these studies interviewed or attempted to interview every one of the members of bounded communities. Table 1 presents characteristics of the datasets. The first three studies come from the diffusion network dataset (Valente, 1995). The oldest study is the 1955 classic Medical Innovation study (Coleman et al. 1966; Burt 1987). Physicians in this study were from four Illinois communities (Peoria, Bloomington, Quincy, and Galesburg) and were asked to name three general practitioners who lived in their communities with whom they discussed medical practices, from whom they sought advice, and whom they considered friends.

Description of Datasets

Dataset	Year Data Collected	Setting	Make up of networks	No. of network questions	Question(s) asked
1	1955	Illinois communities	Physicians	3	Name 3 physicians who you consider friends, with whom you discuss medical practices, & from whom you seek advice
2	1973	Rural villages in Korea	Married women of childbearing age	1	Name 5 people in the village from whom you seek advice about family planning
3	1966	Rural villages in Brazil	Farmers	3	Name 3 best friends, 3 most influential people in community, & 3 most influential farmers
4	1993	Urban Cameroon	Women members of a voluntary organization	1	Name 5 friends belonging to the voluntary organization
5	1993	Urban Cameroon	Women members of a voluntary organization	1	Circle names of all organization members considered friends
6	1991	Corporate law firm ir	All attorneys	3	Circle names of all other attorneys considered strong coworkers, friends & individuals to whom you would go for advice
7	1996	IT department in a company in Latin America	All information technology (IT) employees	7	7 separate questions regarding information exchange at work
8	1996	IT department in a company in the US	All information technology (IT) employees	7	7 separate questions regarding information exchange at work

Data for study two were collected in 1973 in a study of the diffusion of family planning practices in Korea (Rogers & Kincaid 1981). Women in rural villages were asked to nominate five other village residents from whom they sought advice about family planning. Data from the third study were collected in rural villages in 1966 in a study of the spread of farming practices in Brazil. Farmers were asked to name their three best friends, the three most influential people in their community, and the three most influential farmers in their community.

Data for studies four and five were collected in 1993 from women's voluntary associations, Tontines, in urban Cameroon using both nominations and roster data collection techniques (Valente et al. 1997). Study participants initially were asked to nominate five friends who were members of their voluntary organization. In a separate question, study participants were asked to circle the names of friends on a roster, which listed the names of all members of the voluntary organization. These two questions may generate different networks and therefore were considered as two distinct datasets and centrality measures are calculated for each separately.

In these first five studies, network data were collected to study the spread of a new idea, opinion, or practice (Valente 1995; Rogers 2003). In the last three studies, network data were collected in order to assist executives in organizations to better understand the flow of information within and between organizations. Data for study six were collected in 1991 from all the attorneys, partners, and associates, employed in a law firm (Lazega & van Duijn 1997). A second distinction is that the boundary for this network was functional rather than geographic. The law firm had multiple offices throughout the U.S. and as such the network data were collected among employees working in offices located in three different U.S. cities. Data for study seven were collected in 1996 from the information technology (IT) personnel within a U.S. company (Krebs, 2002).

In the law firm, attorneys were asked in three separate questions to nominate other lawyers within the firm whom they would consider to be close coworkers, friends, and individuals to whom they went for advice. Attorneys were given a roster of names and were allowed to nominate as many other attorneys from the roster as they chose for each question. In the high tech firm, IT employees were asked seven separate questions regarding the exchange of specific types of work-related information. For each question, they were allowed to select an unlimited number of names from a roster, which listed all other IT personnel employed by their firm.

All of the sociometric networks included in this study differ in their size, the number of questions asked of respondents, the type of questions asked, and the number of nominations allowed. Table 2 summarizes these differences and shows that most of these studies collected data from more than one network. For example, the Brazilian farmer's study interviewed farmers living in 11 different villages. The total number of networks in these 7 studies is 62.

Given that our aim was to determine how well centrality measures correlated with one another, we felt it would be more difficult to make this comparison if information from a large portion of the network was not collected. Therefore, we excluded from our study any network in which less than 50% of the enumerated population initially responded to the network questions. Using this criterion, we excluded one of the Illinois communities, one Korean village, and one of the Cameroonian women's voluntary organizations, leaving a final sample of 58 networks. (Since the roster data and the nominations data for the Cameroonian women's voluntary organizations were considered as two distinct datasets, exclusion of data from one of the women's voluntary organizations resulted in the loss of two networks). Table 2 presents the average properties of the networks in the 7 studies. Since networks in the same study often shared similar attributes, it would be cumbersome to present the characteristics of all 58 of these networks. . Further information on these datasets are available in Costenbader & Valente, 2003; Valente, 1995.

Table 2. Network Characteristics (N=58).

Dataset	Number of networks analyzed *	Average network size	Average response rate	Average network density	Total number of nominations possible	Average number of nominations	Range of Out-degree nominations sent	Average network centralization (symmetrized)	Average network centralization (in-degree)	Average network centralization (out-degree)
1	3	64	56%	0.06	9	2.61	0-8	24.11%	20.04%	12.26%
2	24	68	64%	0.03	5	1.64	0-5	20.02%	21.06%	5.12%
3	11	76	82%	0.03	9	1.94	0-7	27.35%	30.04%	5.77%
4	9	83	76%	0.04	5	3.13	0-5	22.08%	28.65%	2.03%
5	9	83	76%	0.49	unlimited	39.06	0-152	28.82%	16.77%	49.77%
6	1	71	100%	0.32	unlimited	22.15	2-49	33.23%	30.64%	39.46%
7	1	72	82%	0.20	unlimited	14.19	0-34	24.39%	24.34%	28.69%
8	1	45	96%	0.38	unlimited	16.62	0-40	43.45%	35.74%	54.34%

* Networks in which the response rate was less than 50% were excluded from our analysis.

RESULTS

Table 3 reports the average correlations among the measures. The overall correlation among all 9 measures and 58 datasets was 0.53 (SD=0.14). Correlations among specific centrality measures varied. For example, degree (symmetrized) had the strongest overall correlations at 0.70 with about the same standard deviation (SD=0.15). Eigenvector centrality had the next highest average correlation (r=0.67, SD=0.15) and in-degree, out-degree, betweenness, and symmetrized closeness all had similar correlations (average r=0.53 to 0.58). Directional closeness measures, in-closeness and out-closeness, had the lowest average correlation (0.34 and 0.44, respectively).

Table 3. Average correlations between centrality measures (N=58).

		1	2	3	4	5	6	7	8	9	10	11	
1	Indegree												
2	Outdegree	0.3											
3	Degree	0.78	0.71										
4	Between	0.62	0.54	0.7									
5	S-Between	0.69	0.5	0.85	0.67								
6	Closeness-In	0.55	0.16	0.45	0.37	0.31							
7	Closeness-Out	0.18	0.81	0.56	0.39	0.38	0.02						
8	S-Closeness	0.4	0.64	0.66	0.37	0.44	0.42	0.65					
9	Integration	0.7	0.26	0.58	0.5	0.41	0.9	0.15	0.51				
10	Radiality	0.21	0.86	0.61	0.44	0.41	0.06	0.98	0.67	0.19			
11	S-Int/Rad	0.45	0.7	0.73	0.43	0.5	0.44	0.69	0.99	0.54	0.72		
12	Eigenvector	0.71	0.69	0.92	0.64	0.72	0.44	0.55	0.63	0.57	0.59	0.71	
Average		0.51	0.56	0.69	0.52	0.53	0.37	0.49	0.58	0.48	0.52	0.63	0.65
Standard Deviation		0.21	0.23	0.14	0.16	0.14	0.27	0.22	0.25	0.28	0.17	0.12	0.12

The correlations between measures were also quite varied. The highest correlation was between eigenvector centrality and degree (average r=0.92), perhaps because both measures are symmetrized and rely, to some extent, on direct connections. The next highest correlation was between symmetrized betweenness and degree (average r=0.85) followed by closeness-out and out-degree (average r=0.81).

The correlation between in-degree and degree is considerably higher than the correlation between out-degree and degree. In part this reflects the nature of the data analyzed in this study. Since most of these datasets involve a limited number of sociometric choices, there is comparatively less variation in out-degree than in-degree. And since degree is calculated on both the row and column sums of the adjacency matrix, it will correlate more strongly with in-degree.

The lowest correlation between measures was between closeness-out and closeness-in (average r=0.01). This is surprising, indicating that the direction of the calculation matters more than the property being measured by the algorithm. It is also worth noting that the standard deviation of these correlations was highest (SD=0.39). The next lowest correlations were for closeness-in and out-degree (average r=0.16) and closeness-out and in-degree (average r=0.18). This is most likely a consequence of variation in naming network partners. Someone who named only 1 person, but who was named by many others would have high closeness-in but very low out-degree.

Not surprisingly, eigenvector centrality is more strongly correlated with the symmetrized versions of the other measures than with their asymmetric versions. For example, the average correlation with degree was 0.91 whereas it was 0.71 and 0.69 for in- and out-degree.

By study. There is some variability in the overall correlation among measures between studies. Table 4 reports the total average correlation between studies and shows that the IT department and Lawyers studies had the highest average overall correlations (average r=0.89 and r=0.82, respectively). In contrast the medical innovation and Cameroon roster had substantially lower average correlations (average r=0.44 and r=0.45, respectively). Restricting the comparison to studies with multiple networks yielded a more statistically significant difference in averages between studies. The Bonferroni test showed the difference was primarily between the low average correlation within the Cameron roster data and the somewhat higher ones in the Korean Family Planning and Brazilian Farmers data. Restricting the comparison to the 2 Cameroon studies that used the same questions and populations yielded a marginally non-significant difference in average correlation, and a significant difference in variance (Bartlett's chi-square test for equal variances, χ^2 =4.45, df=1, p<.05).

Table 4. Total average correlation between studies (N=58).

Dataset	Total Correlation	
Medical Innovation	0.43	
Korean Family Planning	0.57	0.57
Brazilian Farmers	0.54	0.54
Cameroon Nominations	0.53	0.53
Cameroon Roster	0.46	0.46
Lawyers	0.81	
IT Dept. Latin America	0.89	
Total	0.54	0.54
	p<.05	p<.01

Network properties. We tested whether network properties (e.g., density, reciprocity, centralization) affected the correlation among measures. Reciprocity was strongly associated with centrality measure correlations (β=0.89, p<.01). If there were many reciprocated relationships in the network, the various centrality measures were highly correlated. This strong correlation could be a function of the symmetry status of the various measures – networks with higher levels of reciprocity will have higher correlations between asymmetric measures than those with lower levels of reciprocity. For example, the correlation between in-degree and out-degree will be unity when the network is perfectly symmetric because the in- and out-ties are identical.

Table 5. Average correlations classified by symmetry status of calculation regressed on study and network properties.

	All Measures	Symmetric w/ Symmetric	Symmetric w/ Asymmetric	Asymmetric w/ Asymmetric
Average Correlation	0.54	0.71	0.44	0.55
Cameroon Roster	-0.11	0.06	-0.16	-0.11
Brazilian Farmers	0.20	-0.04	0.16	0.23*
Korean Family Planning	0.31*	0.01	0.29*	0.31**
Density	-0.52	0.67**	-0.62*	-0.55
Reciprocity	0.85**	0.26**	0.68**	0.83**
Degree Centralization	0.02	-0.07	0.07	-0.02

As table 5 shows, although reciprocity was always positively associated with average measure correlation, the association between reciprocity and centrality measure correlation varies by the symmetry status of the measures. Correlations between symmetrized measures are weakly associated with reciprocity (β=0.29, p<.01), those between symmetric and asymmetric measures very strongly associated with reciprocity (β=0.84, p<.01), and those between asymmetric measures also strongly associated with reciprocity (β=0.69, p<.01). The average correlation between asymmetric centrality measures increases with the reciprocity of the network. Measures based on symmetrized data are unaffected by the degree of reciprocity in the network because the reciprocity has already been forced.

Density is also associated with the average correlation. Density has a negative but not statistically significant association with the average correlation of all measures (β=-0.25, p=NS). The correlation is positive and significant for symmetric measures (β=0.60, p<.01) and negative and significant for symmetric measures with asymmetric ones (β=-0.60, p<.01). This demonstrates that the density of a network plays a role in how well different centrality measures correlate to one another. That is, symmetrizing data in low density networks adds links to the network, which changes the centrality calculations so that the symmetric and asymmetric versions diverge. Interestingly, degree centralization does not affect the correlation between measures.

Finally, the number of components in the networks was positively associated with centrality correlations among asymmetric measures only (β=0.28, p<.05). Networks with more components had stronger correlations among asymmetric centrality measures. Some caution in interpretation of these regression results is warranted as many of these structural properties are correlated. For example, density is positively correlated with reciprocity (β=0.51, p<.01) and centralization (β=0.37, p<.01); and negatively correlated with number of components (β=-0.60, p<.01). Reciprocity is negatively correlated with number of components (β=-0.53, p<.01).

DISCUSSION

We find strong but varied correlations among the 9 centrality measures presented here. The average of the average correlations was 0.53 with a standard deviation of 0.14, indicating that most correlations would be considered strong. The level of correlation among measures seems nearly optimal - too high a correlation would indicate redundancy and too low, an indication that the variables measured different things. The amount of correlation between degree, betweenness, closeness, and eigenvector indicates that these measures are distinct, yet conceptually related.

Direction matters, as the correlations for the symmetrized measures were quite different than those for the asymmetric versions. Interestingly, the only network variable that was positively and significantly associated with correlations between all centrality measures (correlations between symmetric measures, asymmetric measures and between symmetric and asymmetric centrality measures) was reciprocity, suggesting that the more bi-directional information flow between individuals is, the less distinct centrality measures become. In addition, correlations between symmetrized measures were associated with the number of components and network density, while asymmetric measures were not. Thus, symmetrizing matrices before making centrality calculations should be done with caution and only if justifiable substantively. In addition, unsymmetrized centrality measures might be more distinct in densely connected networks with more components.

Correlations also varied by study, but not in obvious ways. The Lawyers and IT department studies had the highest overall correlations, perhaps because these studies used roster methods and were considerably denser than the other studies, which used nomination methods. The Cameroon roster study, however, had the second lowest average correlation and it was significantly lower than the Cameroon nomination study. Thus, network data collection methods appear to influence the correlation between centrality measures and therefore should be considered when comparing centrality measures across studies.

Network properties such as reciprocity and density correlated with average correlation in interesting ways. Density decreased the correspondence between centrality measures when comparing symmetric measures with asymmetric ones. Density increased the correspondence between centrality measures when those measures were calculated on symmetric data. In contrast, in networks with many reciprocated relationships, centrality measures calculated on symmetric data provided more unique information than those calculated on asymmetric data Perhaps this occurred because the reciprocal nature of relationships decreased the differences in centrality measures. Overall, our findings show that symmetrizing network data creates disparities between symmetric and asymmetric centrality measures.

References

Becker, M. H. (1970). Sociometric location and innovativeness: Reformulation and extension of the diffusion model. American Sociological Review, 35, 267-282

Bolland, J.M. (1998). Sorting Out Centrality: An analysis of the Performance of Four Centrality Models in real and simulated networks, Social Networks, 10, 233-253.

Bonacich, P. (1972). Technique for Analyzing Overlapping Memberships, Sociological Methodology, 4, 176-185.

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

Borgatti, S.P. (2005). Centrality and Network Flow. Social Networks. 27, 55-71.

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

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

Costenbader, E., & Valente T.W. (2003). The stability of centrality measures when networks are sampled. Social Networks, 25, 283-307.

Degenne, A. & Forsé, M. (1999). Introducing social networks. Sage: Thousand Oaks, CA.

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

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

Freidkin (1991). Theoretical foundations for centrality measures. The American Journal of Sociology, 96(6), 1478-1504.

Galaskiewicz, J. (1985). Social organization of an urban grants economy. New York: Academic Press.

Krebs, V. http://www.orgnet.com/IHRIM.html . Accessed February 2002.

Lazega, E. and van Duijn, M. (1997). Position in formal structure, personal characteristics and choices of advisors in a law firm: a logistic regression model for dyadic network data. Social Networks, 19, 375-397.

Newman, M.E.J. (2003). Ego-centered networks and the ripple effect. Social Networks, 25, 83-95.

Rogers, E. M. (2003). Diffusion of Innovations (5th ed.). New York: The Free Press.

Rogers, E. M., & Kincaid, D. L. (1981). Communication Networks: A new paradigm for research. New York: Free Press.

Rothenberg, R. Potterat, J.J., Woodhouse, D.E., Darrow, W.W., Muth, S.Q., & Klovdahl, A.S. (1995). Choosing a centrality measure: Epidemiologic correlates in the Colorado Springs study of social networks. Social Networks, 17,273-297.

Sampson, S. (1969). Crisis in a cloister. Unpublished doctoral dissertation. Cornell University.

Scott, J. (2000). Social network analysis: A handbook (2nd Ed.). Newbury Park, CA: Sage.

Valente, T. W. (2005). Models and methods for innovation diffusion. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.) Models and Methods in Social Network Analysis. Cambridge, UK: Cambridge University Press.

Valente, T. W., Watkins, S. C., Jato, M. N., et al., (1997). Social network associations with contraceptive use among Cameroonian women in voluntary associations. Social Science & Medicine, 45, 677-687.

Valente, T. W. (1995). Network models of the diffusion of innovations. Cresskill, NJ: Hampton Press.

Valente, T. W., & Davis, R. L. (1999). Accelerating the diffusion of innovations using opinion leaders. The Annals of the American Academy of the Political and Social Sciences, 566, 55-67.

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

Wasserman, S., & Faust, K. (1994). Social Networks Analysis: Methods and Applications. Cambridge, UK: Cambridge University Press.

Analysis of Transitivity and Reciprocity in Online Distance Learning Networks[1]

Reuven Aviv
Department of Computer Science, Tel Hai Academic College and the Open University of Israel

Zippy Erlich
Department of Computer Science, Open University of Israel

Gilad Ravid
Annenberg School of Communication, University of Southern California

The goal of this research is to explore whether network structures common in social networks also emerge in online distance learning networks. We compare the observed values of reciprocity and transitivity in 95 online distance learning networks and 40 social networks with predictions of Random Graph models. All the networks tested exhibit reciprocities that significantly deviate from the predictions of the random graph models. All the online distance learning networks and some, but not all, of the social networks exhibit transitivity compatible with generalized random graph models. We provide possible explanations for these behaviors, based on the broadcast nature of the online distance learning networks, and practical implications for online collaborative learning.

Corresponding author: Prof. Reuven Aviv, Department of Computer Science, Tel Hai Academic College, Upper Galilee 12210, Israel. Email: reuvenav@telhai.ac.il.

We thank the anonymous referees for invaluable comments that helped improve the first draft of this paper. We also thank Gila Haimovic for critical review of the manuscript.

INTRODUCTION

Online learning envisions the Internet primarily as a communication facilitator among all parties involved, and secondarily as a medium for distribution of educational materials (Mayadas 2000). Thus, the underlying assumption is that learners are distributed collaborating active agents who are purposefully seeking and constructing knowledge within a meaningful context (Harasim 1990; Harasim et al. 1995; Hiltz 1994), thereby developing their cognitive skills (Oshima, Bereiter, and Scardamalia 1995). For a summarizing review, see Hsiao (2000). A recent comprehensive compendium can be found in Anderson and Elloumi (2004).

The learning community typically uses an online distance learning network for its communication. What are the observed characteristics of these networks? Do they maintain various types of communication like small village networks, or are they single-purpose, like a trigger response gene network (Milo et al. 2002)? Do they have a distributed or centralized power of influence? Are they cohesive? Some of these questions were studied by analyzing topological features of online distance learning networks: power of influence (Martinez et al. 2002), correlation of power distributions (Cho, Stefanone, and Gay 2002), evolution of cohesion (Reffay and Chanier 2002), the relations between cohesion, roles, and knowledge construction (Aviv et al. 2003), and time variation and media dependence of communication patterns (Haythornthwaite et al. 2000).

Crucial for collaboration in the learning community is the development of responsiveness among the participants (Rafaeli et al. 1998). This relation is the cohesive glue between the actors in the learning community. What are our theoretical expectations about this relation? Numerous reports emphasize the social nature of online distance learning networks (Richardson and Swan 2003; Wegerif 1998; Haythornthwaite 2002; Brett and Nagra 2005). Hence, we focus our study on well known attributes of social networks (Monge and Contractor 2003; Contractor, Wasserman, and Faust 1999).

Reciprocity is suggested as one of the defining attributes of any network, real life (Seabright 2004) or virtual (Wellman and Gulia 1999). The underlying assumption is that an actor forges relations with someone who has already related to him or her, or with someone who is a promising resource and will probably reciprocate. In fact, this is a common attribute of social networks of people and animals (Skvoretz 2002), and of organizations (Monge and Contractor 2001). Generalizing, the theory of distributed learning (Dede 1996) posits that online learners construct knowledge via reciprocal growth of cognitions of members of the learning group, and reciprocity is incorporated as one of the "principles" of good practice in education (Chickering and Gamson 1987) or as a behavioral indicator for the emergence of a network (Herring 2001). One empirical study, Hakkinen, Jarvela, and Byman (2001), identified reciprocity of interaction in online networks, and another, Wang and Fesenmaier (2003), found that reciprocation is one of the major motivations driving individuals' contributions in online networks.

It could be argued that reciprocity cannot be easily developed in online distance learning networks. To develop reciprocity, learners have to go through a process of assessment of risks, rewards, and the likelihood of reciprocation, as well as trust development. Simulation analysis of the Prisoner's Dilemma game (Axelrod 1990) illustrates this idea: when two actors don't trust that their peers will reciprocate, they initially adopt the less risky, lower benefit strategy of non-reciprocating; then they may reach a reciprocation state via a series of rounds in which they learn the strategies of their partners. Psychology suggests that during the learning period, each of the actors develops three entities (Aron 1996): ego, other, and the reflective-self, which is awareness of ego as the object of one's own investigation as well as the object of investigations by others (Kadushin 2004). It is feasible to establish such a learning period in social networks by using pre-existing, rich "wide-bandwidth" social links. But relations between actors in online distance learning networks are not necessarily rich. In many cases, students rarely meet, and the existence of the network is limited in scope and in time. The social links have a narrow bandwidth and this, it seems, reduces the likelihood of reciprocation and makes it very difficult to implement a learning period in online distance learning networks.

Transitivity is another common attribute of social and organizational networks, as well as of many technological and biological networks (Newman 2003). The underlying assumption is that a transitive "cognition balance" mechanism is developed in social communities aiming at overcoming dissonance and consistency in cognition among actors (Heider 1958; Festinger 1957; Cartwright and Harary 1956). For example, when dissonances arise between people, they attempt to reduce them by persuading others, who will persuade more people, and so on. To date, transitivity has not been incorporated into models of online distance learning networks. Indeed, it could be argued that in these networks, transitivity will not be developed. In distance learning networks, each of the participants is interested in a certain small issue at a certain point in time (usually related to a specific assignment), whose scope is, in many cases, limited – it is of interest to few actors. Other issues, or even related concepts that have no direct relation to the query, are of less significance to the actor involved in the discussion, let alone to other actors. The lifetime of the issue is short (usually until some due date). Thus, there is no group-wise drive to settle conceptual inconsistencies regarding past issues, or dissonances in perceptions regarding others. Therefore, no transitive cognition balance mechanism is needed and none, presumably, will be established.

As a result, our starting point in this research is to hypothesize that in contradistinction to social networks, neither reciprocity nor transitivity develops in online distance learning networks. We will test these hypotheses indirectly by statistical comparative analysis of the reciprocity and transitivity of a large set of online distance learning networks and social networks in relation to several random graph models. The precise formulation of the hypotheses is described in the next section. In subsequent sections, we describe the database used in this study, the results, and the significance of the results is discussed in the last section.

Research Questions

We consider an online distance learning network to be one that employs text-based communication. It is a broadcast network: any posted message is readable by all members (called "actors" in this study). The major expectation of actors in such networks is that their messages will be responded to (Rafaeli 1988; Rafaeli and LaRose 1993), but this does not always happen. Thus, response links might or might not develop: A response link is defined to be realized from actor i to actor j if the number of messages posted by i and responded to by j is above threshold, defined in this study as 1. We call the collection of the actors and the response links the response graph of the online distance learning network. The response graph is akin to a particular relation in a social network.

The state of a single actor in the graph is characterized by its actor configuration. This is a set of two integers (d*out*, d*in*): the out-degree, d*out*, counts the number of actors with whom that actor created links; the in-degree, d*in*, counts the number of actors who created links with the focus actor. In the response graph, the degrees can be thought of the response capacities of the actors. In a social network relation, the in-degree is sometimes called the popularity of the actor, and the out-degree, the influence of the actor. The set of all actor configurations is called the degree sequence of the graph.

The state of a pair of actors is characterized by its dyad configuration which is reciprocal (mutual), asymmetric, or null: if the actors have a bidirectional link (e.g., in an online distance learning network they responded to each other at least once) then the pair has realized a reciprocal, or mutual, configuration; if the actors only have a one-directional link, they have realized the asymmetric configuration; otherwise there is no link between them, so they have realized the null configuration. In this research we use the number of reciprocal configurations in a graph as a measure of its reciprocity, although other definitions could also be used (Rao and Bandyopadhyay 1987). The number of reciprocal (mutual), asymmetric, and null configurations, denoted by M, A, and N respectively, is called the dyad census of the graph.

A triplet of actors can have several configurations (see Wasserman and Faust (1994) for details). For the purpose of this research, we are interested in just one. A triplet (i, j, k) realizes a transitive configuration if they are connected by the three response links i→j, j→k, and i→k. In the response graph actor i responded to j, j responded to k, and i also responded to k. Note that these responses need not be sequential. We use the number of transitive configurations in the graph as a measure of its transitivity,

We wish to reveal the structures emerging in online distance learning networks and social networks, above and beyond random interactions. To do this we compare reciprocity and transitivity of observed online distance learning networks or social networks with predictions of Random Graph models (Wasserman and Faust 1994) with various degrees of constraints.
Each of these models creates an ensemble of graphs; the creation is controlled by the probability distribution function of the model; i.e., the probability assigned to the creation of each of the graphs in the ensemble. The observed graph is tested for the likelihood of its occurrence in the ensemble. We use four classic models:

1. The Erdös-Rényi model (Erdos and Renyi 1960): Graphs are generated at random, where each link has a certain link-probability, p, to exist. The probability distribution function is binomial with

parameters N(N-1) and p, where N is the number of actors.

2. The Holland-Leinhardt model (Holland and Leinhardt 1976): The probability distribution function is uniform, conditioned on the dyad census. This function is usually denoted by U|(M,A,N).

3. The Molloy-Reed model (Molloy and Reed 1995): The probability distribution function is uniform, conditioned on the in-degrees and the out-degrees sequences. This function is usually denoted by U|({Xi+}, {X+i}), where Xi+, X+i denote the out-degree and the in-degree sequences, respectively.

4. The Snijders model (Snijders 1991): The probability distribution function is uniform, conditioned on the dyad census, the in-degrees and the out-degrees sequences. This function is usually denoted by U|(M, {Xi+}, {X+i}).

Formally, we will test the following five hypotheses for each of the observed online distance learning networks:
H1: the observed reciprocity can be explained by the Erdös-Rényi model

H2: the observed reciprocity can be explained by the Molloy-Reed model

H3: the observed transitivity can be explained by the Holland-Leinhardt model.

H4: the observed transitivity can be explained by the Molloy-Reed model.

H5: the observed transitivity can be explained by the Snijders model.

The rationale for choosing these particular models is that they impose increasing levels of constraints on the otherwise random behavior of the actors. The Erdös-Rényi model imposes no constraints; the Holland-Leinhardt model imposes fixed reciprocity; the Molloy-Reed model imposes fixed degrees; and the Snijders model imposes both constraints. Moreover, the Erdös-Rényi and Hollad-Leinhardt models assume no correlation between the probabilities for creating links or dyads, respectively, whereas such correlations are inherent in the generalized random graph models of Molloy-Reed and Snijders. We include at least one of each of these types of models in analyzing each of the sub-structures, as they could provide hints to the origin of excessive sub-structures, if this exists. The Erdös-Rényi and the Molloy-Reed models do not carry built-in reciprocity, so we use them as base-line models for identifying reciprocity beyond random links. The Holland-Leinhardt and the Molloy-Reed models do not carry built-in transitivity, so we use these models to identify transitivity beyond randomness. In addition, we use the Snijders model because a comparison of observed transitivity with the values of both the Molloy-Reed and the Snijders models will tell us whether excessive transitivity (if it exists) is an artifact of reciprocity or not.

METHODS

The Open University of Israel is a distance learning institution, based heavily on intensive use of its learning technologies environment, with optional face-to-face tutorials. Each course utilizes at least one online distance learning network, usually over one semester (16 weeks). The objectives of the networks vary – from collaborative knowledge construction (Aviv et al. 2003), to social, pedagogical or technical support. Objectives are not mutually exclusive. Accordingly, size, response links and participation patterns vary from course to course and from semester to semester. Numbers of participants in the network vary from 10 to 150, but most have about 50 students.

In this study, we selected for analysis a sample of 95 of about 500 online distance learning networks in the Open University of Israel. Networks included in the sample were selected at random, omitting 5 networks in which the number of active participants (those who post at least one message during the semester) was below an arbitrarily selected threshold of 10.

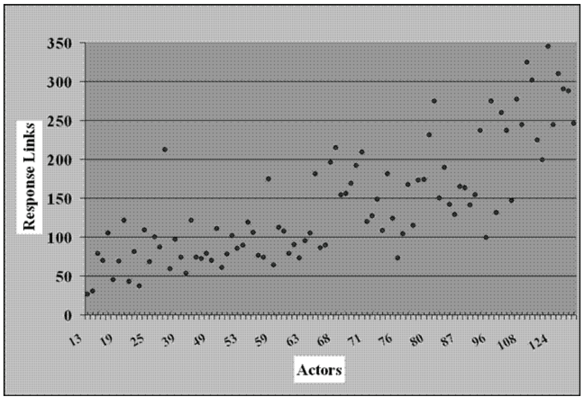


Figure 1

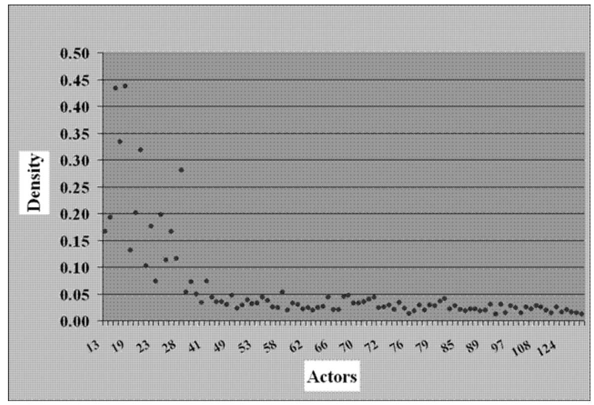


Figure 2

For each of these networks we created its (observed) response graph. The global characteristics of the observed response graphs are presented in Figures 1 and 2. In these figures, as in all other figures in this paper, each point represents one graph (or network). Figure 1 shows that the number of nodes (actors), N, ranges from 13 to 140, and the number of response links, L, ranges from 30 to 350. Increasing the number of participants increased the number of posted messages, which in turn, increased the number of responses to posted messages. Figure 2 shows that the density, L/[N(N-1)] decreases from 0.02 to 0.45 . Most networks are sparse.

For comparative purposes, we also analyzed a set of 40 well-known social networks (Kapferer 1972; Knoke and Kuklinski 1982; Krackhardt 1987; Wasserman and Faust 1994), capturing friendship, advice seeking, assistantship, message exchange, and trade relations. The characteristics of these networks are presented in Figure 3. The range of links and densities are similar to those of the response graphs. Note, though, that the number of nodes in these networks (not shown) was limited to the narrow range 21 – 39, which explains the straight line relation between the density and the number of links.

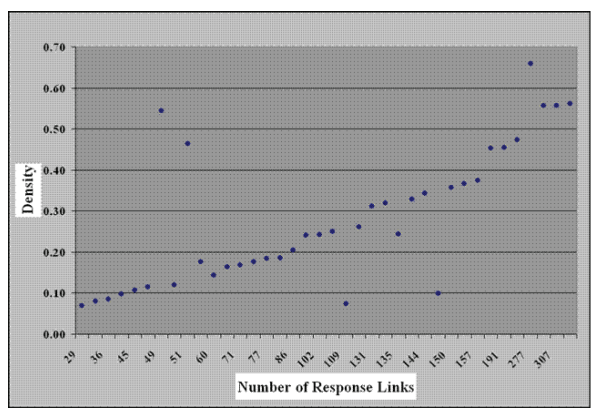


Figure 3

Each response graph is represented by a binary adjacency matrix: rows and columns are labeled by the nodes; an entry (i, j) is 1 if a response link exists from i to j. Otherwise it is 0. The adjacency matrices of the observed response graphs were constructed by a computer program, opus2sna, developed in-house, that scanned the recorded transcripts of the messages sent in the networks, and identified posters and responders. The adjacency matrices of the social networks were drawn from the cited references. Once the adjacency matrix was known, we calculated the reciprocity and the transitivity following (Wasserman and Faust 1994).

In the Erdös-Rényi model, dyads are independent Bernoulli variables with probability p^2, where the link probability, p, is estimated by the observed density of the graph. The expected number of reciprocal dyads is then $0.5N(N-1)\,p^2$. Expected values in other models were estimated by the network simulation program Netminer (NetMiner CYRAM Co. Ltd 2004). For each pair of observed graphs and for each hypothesis, we ran the random graph model referred to in the hypothesis to simulate an ensemble of 100,000 graphs. With this number of simulations, the results were stable. The constraints (M, A, N and the degree sequences) were taken as the observed values. The program provided the expected values and the fraction of ensemble graphs with reciprocity and/or transitivity above and/or below the observed value. If either of these fractions is smaller than the significance level of 0.01, the hypothesis was rejected for that observed response graph. Otherwise, the hypothesis was accepted.

RESULTS

Figure 4 presents for each of the 95 online distance learning networks, the observed and the expected values of reciprocity, according to the Erdös-Rényi and the Molloy-Reed models.

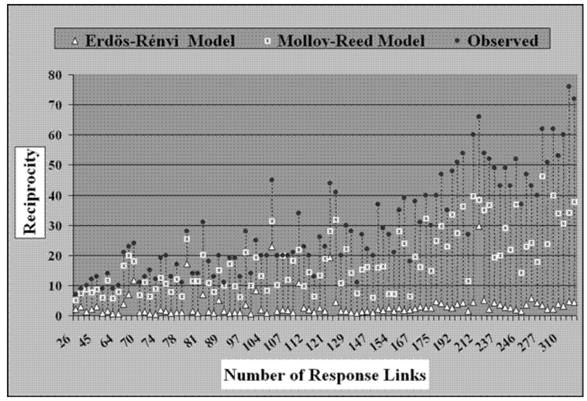


Figure 4

Figure 5 presents the same information for the social networks. In all cases, the observed values are substantially larger than the values expected by the two models; specifically, for each of the observed online distance learning networks and social networks, the probability that the models will generate reciprocity equal or larger than the observed values is smaller than p = 1%. Hypotheses H1 and H2 are rejected for all the online distance learning networks and social networks.

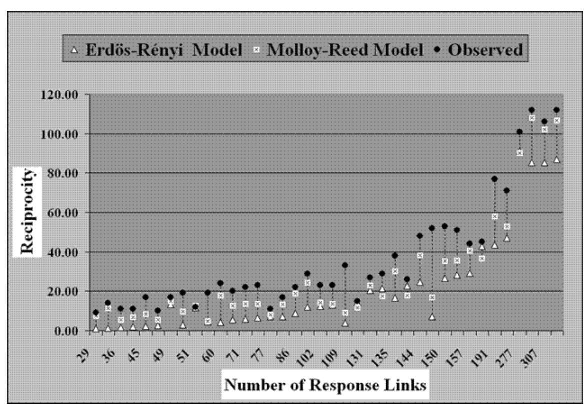


Figure 5

Figure 6 presents, for each of the 95 online distance learning networks, the observed and predicted values of transitivity according to the Holland-Leinhardt, Molloy-Reed and Snijders models. The results for the social networks are presented in Fig. 7. In all these cases, the observed values are substantially higher than the values expected by the Holland-Leinhardt model, at a level of $p < 0.01$. In the online distance learning networks, the observed transitivity values agree with the values expected by the Molloy-Reed and the Snijders models (which agree with each other); the probability that these models will generate networks with transitivity equal or larger than the observed values is larger than $p = 10\%$.

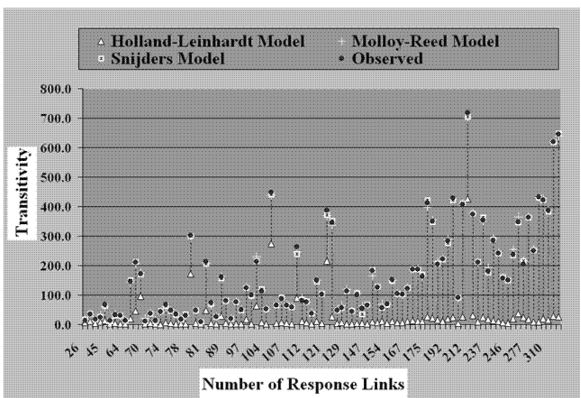


Figure 6

The behavior of the social networks is less homogeneous: In many social networks, the transitivity agrees with the value expected by the Molloy-Reed and the Snijders models (which again agree with each other), but in other social networks, the transitivity deviates significantly (in the statistical sense) from the expected predictions (note the logarithmic scale in Figure 7). Hypothesis H3 is rejected for all the online learning networks and social networks. Hypotheses H4 and H5 are accepted for all the online learning networks and most – but not all – of the social networks; they are rejected for some social networks.

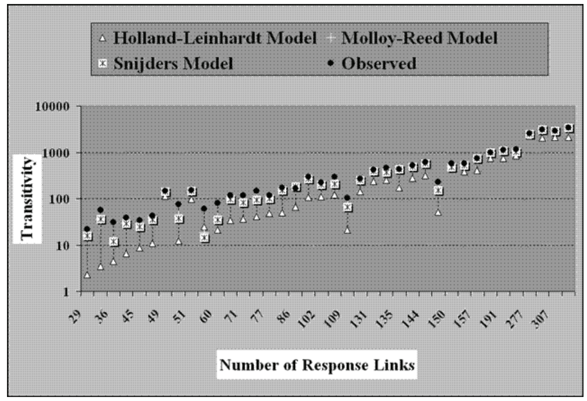


Figure 7

DISCUSSION

The predictions of the simple random graph models (Erdös-Rényi and Holland-Leinhardt) do not fit – in all the tested networks – the observed values of reciprocity and transitivity, respectively. The reason is that the models assume independence of links and dyads, respectively. Most of the tested networks are sparse. Link or dyad independence in such networks predicts low values of the higher sub-structures, reciprocal dyads or transitive triads. This failure is similar to the inadequacy of the Erdös-Rényi model in describing many biological and technical networks (Newman 2003). Links and dyads in nature are typically correlated. This observation led researchers to adopt the generalized random graphs, specifically the Molloy-Reed model, as the basis for analyzing the structures, as well as the dynamics, of naturally occurring and artificial networks (Newman, Strogatz, and Watts 2001). Sub-structures occurring at rates, which are significantly larger than the predictions of this model, are considered "motifs" that require specific explanatory mechanisms (Milo et al. 2002).

The results of this research indicate that social networks and online distance learning networks have similar but not identical motifs. Some – but not all – of the social networks, and all of the online distance learning networks, exhibit transitivity, which is compatible with the expectation of the Molloy-Reed model; transitive configurations are not motifs in these networks. In all the social networks and online distance learning networks, the observed reciprocity significantly exceeds the Molloy-Reed expectations; reciprocal configurations are motifs in all the networks. This points to the similarities and differences in the underlying mechanisms. While there is evidence for a cognitive balance mechanism in at least some of the social networks, as was previously observed in numerous studies (Wasserman and Faust 1994), there is no evidence for this mechanism in all the online distance learning networks. On the other hand, there is evidence for the exchange mechanism in all networks, social and online.

Note that in some social networks there was an excess of transitivity, relative to the predictions of both the Molloy-Reed and the Snijders models (which agree with each other). This means that in these networks, transitivity is a real motif – it is not an artifact of excess in reciprocity; a genuine cognition balance process was at work in these social networks.

As noted in section 1, the lack of a cognitive balance mechanism in typical online distance learning networks is not surprising. As there is no group-wise drive to build a consensus, no transitive cognition balance mechanism is needed. But transitive structures are building blocks of more complex cohesive structures such as response-cliques, which facilitate constructing knowledge by consensus (Aviv, Erlich, and Ravid 2004). If the goal of the online distance learning network is indeed knowledge construction, then suitable collaborative features – i.e. positive interdependence (Johnson and Johnson 1992, 1999) – should be designed in order to initiate the cognition balance mechanism. Indeed, in a comparative case study analysis, Aviv, Erlich, and Ravid (2005) found significant transitivity in one special online distance learning network that was designed and monitored to work as a team to reach a consensual goal (submission of a joint proposal); no such transitivity was found in the more common Q&A-type network. It seems that the goal-directed design of the team network forced its participants to reach consensus, which led to the cognition balance mechanism. This intriguing idea should be further explored on a much larger scale.

Having established the existence of a reciprocal exchange mechanism, we now turn to the question of its emergence in an online distance learning network. As noted in section 1, to establish reciprocity, the actors need to go through a learning period during which they develop the three psychological components of a reciprocal dyad. How can that happen in narrow-band online distance learning networks? The key explanation is that the networks use a broadcast communication mechanism – posted messages readable by all – so actors learn relatively quickly who is, and who is not, a potential reciprocator. The learning period is thus shortened. Moreover, actors develop their ego via their postings: postings are their "public appearances"; they exhibit their own behavioral aspects, such as providers of support or technical advice

(Walther 1994; Hiltz, Johnson, and Turoff 1986; Constant, Sproull, and Kiesler 1996), and they attract respect from others. In addition, actors realize that they have to contribute, and possibly to reciprocate, to get anything at all. This leads to the development of the reflective-self component and to the awareness of the other. Finally, interaction and reciprocation are facilitated by the current online communication environment. All these considerations lead (Wellman and Gulia 1999) to the conclusion that reciprocity is indeed feasible in online networks. This conclusion is supported by the findings of this research.

Reciprocity does not come without a price. The personal viewpoint of an actor in a broadcast environment is that the most efficient way to gain social capital is to do nothing. Thus, the basic tendency of actors is not to respond at all, let alone reciprocate. Certain design features must be in place to provide responsiveness. The usual procedure is to assign the role of major responder to one of the actors, usually a tutor. This leads to the most common type of online distance learning network – the Q&A network. Reciprocity in this case implies that students prefer to respond to the tutor than to their peers. If interaction between peers is required, we have to distribute the roles of responders among a set of actors.

It should be noted that Open University online distance learners have some, though limited, amount of face-to-face communication. One can argue that this may be a factor in the development of reciprocity and inhibition of transitivity. This question will be dealt with in future research.

This research focused on reciprocity and transitivity. These are just two network characteristics of online distance learning networks. There are many others – clustering, degree and power distribution, and cliquishness, to name a few. Some of these features affect the behavior of various types of networks. For example, certain degree distributions (the so-called “scale-free" distributions (Barabasi and Hawoong 1999)) characterize a large number of extremely large networks (Newman, Barabasi, and Watts 2003; Dorogovtsev and Mendes 2004). "Small - World" topology (regular connectivity with few short-cuts) lead to synchronization between nodes (Barahona and Pecora 2002). These network effects are all interdependent, and can be incorporated into a more general analysis using parametric models, such as p* (Anderson, Wasserman, and Crouch 1999; Wasserman and Pattison 1996), biased net models (Skvoretz 1990), or discriminative classifiers (Middendorf et al. 2004). Such comparative global analyses are required in order to answer the fascinating question: What type of networks are online distance learning networks?

References

Anderson, C. J., S. Wasserman, and B. Crouch. 1999. A p* Primer: Logit Model for Social Networks. *Social Networks* 21:37-66.

Anderson, T., and F. Elloumi, eds. 2004. *Theory and Practice of Online Learning*. Athabasca, AB: Athabasca University.

Aron, L. 1996. *A Meeting of Minds: Mutuality in Psychoanalysis*. Hillside, NJ: Analytic Press.

Aviv, R., Z. Erlich, and G. Ravid. 2004. Mechanisms and Architectures of Online Learning Communities. Paper read at The 4th IEEE International Conference on Advanced Learning Technologies (ICALT), Aug. 30 - Sept. 1, 2004, at Jonesuu, Finland.

Aviv, R., Z. Erlich, and G. Ravid. 2005. Response Neighborhoods in Online Learning Networks: A Quantitative Analysis. *Educational Technology & Society* 8 (4).

Aviv, R., Z. Erlich, G. Ravid, and A. Geva. 2003. Network Analysis of Knowledge Construction in Asynchronous Learning Networks. *Journal of Asynchronous Networks* 7 (3):1-23.

Axelrod, R. 1990. *The Evolution of Cooperation*. New York, NY: Basic Books.

Barabasi, A. L., and J. Hawoong. 1999. Mean-Field Theory for Scale-Free Random Networks. *Physica A* 272:173-87.

Barahona, M., and L. M. Pecora. 2002. Synchronization in Small-World Systems. *Physical Review Letters* 89 (054101).

Brett, P., and J. Nagra. 2005. An Investigation into Students' Use of Computer-Based Social Learning Space: Lessons for Facilitating Collaborative Approaches to Learning. *British Journal of Educational Technology* 36 (2):281-92.

Cartwright, D., and F. Harary. 1956. Structural Balance: A Generalization of Heider's Theory. *Psychological Review* 63:277-93.

Chickering, A., and A. Gamson. 1987. *Seven Principles for Good Practice in Undergraduate Education*. Racine, WI: The Johnson Foundation, Inc/Wingspread.

Cho, H., M. Stefanone, and G. Gay. 2002. Social Network Analysis of Information Sharing Networks in a CSCL Community. In *Proceedings of Computer Support for Collaborative Learning (CSCL) 2002 Conference*, edited by G. Stahl, 43-50. Mahwah, NJ: Lawrence Erlbaum.

Constant, D., L. Sproull, and S. Kiesler. 1996. The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice. *Organization Science* 7 (2):119-35.

Contractor, N. S., S. Wasserman, and K. Faust. *Testing Multi-Level, Multi-Theoretical Hypotheses About Networks in 21 Century Organizational Forms: An Analytic Framework and Empirical Example* 1999. http://www.spcomm.uiuc.edu/users/nosh/manuscripts/pstarpaper.html.

Dede, C. 1996. Emerging Technologies and Distributed Learning. *American Journal of Distance Education* 10 (2):4-36.

Dorogovtsev, S. N., and J. F. F. Mendes. 2004. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford: Oxford University Press.

Erdos, P., and A. Renyi. 1960. On the Evolution of Random Graphs. *Publ. Math. Inst. Hung. Acad. Sci.* 5:17–61.

Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row, Preston & Co.

Hakkinen, P., S. Jarvela, and A. Byman. 2001. Sharing and Making Perspectives in Web-Based Conferencing. In *Proceedings of the First European Conference on Computer-Supported Collaborative Learning*, edited by P. Dillenbourg, A. Eurelings and K. Hakkarainen, 285-92. Maastricht: Universiteit Maastricht.

Harasim, L. M. 1990. *On-Line Education: Perspectives on a New Environment*. New York, NY: Praeger.

Harasim, L. M., S. R. Hiltz, L. Teles, and M. Turoff. 1995. *Learning Networks: A Field Guide to Teaching and Learning on-Line*: MIT Press.

Haythornthwaite, C. 2002. Building Social Networks Via Computer Networks: Creating and Sustaining Distributed Learning Communities. In *Building Virtual Communities: Learning and Change in Cyberspace*, edited by K. A. Renninger and W. Shumar, 159-90. Cambridge: Cambridge University Press.

Haythornthwaite, C., M. Kazmer, J. Robins, and S. Shoemaker. 2000. Community Development among Distance Learners: Temporal and Technological Dimensions. *Journal of Computer Mediated Communication* 6 (1). http://www.ascusc.org/jcmc/vol6/issue1/haythornthwaite.html.

Heider, F. 1958. *The Psychology of Interpersonal Relations*. New York, NY: John Wiley.

Herring, S. C. 2001. Computer-Mediated Discourse. In *The Handbook of Discourse Analysis*, edited by D. Schiffrin, D. Tannen and H. Hamilton, 612-34. Oxford: Blackwell.

Hiltz, S. R. 1994. *The Virtual Classroom: Learning without Limits Via Computer Networks*. Norwood, NJ: Ablex.

Hiltz, S. R., K. Johnson, and M. Turoff. 1986. Experiments in Group Decision Making: Communication Process and Outcome in Face-to-Face Versus Computerized Conferences. *Human Communication Research* 13 (2):225-52.

Holland, P. W., and S. Leinhardt. 1976. The Statistical Analysis of Local Structure in Social Networks. In *Sociological Methodology*, edited by D. R. Heise, 1-45. San Francisco, CA: Jossey Bass.

Hsiao, J. W. D. L. *CSCL Theories* 2000. http://www.edb.utexas.edu/csclstudent/Dhsiao/theories.html.

Johnson, D. W., and R. T. Johnson. 1992. *Positive Interdependence: The Heart of Cooperative Learning*. Edina, MN: Interaction.

Johnson, D. W., and R. T. Johnson. 1999. *Learning Together and Alone: Cooperative, Competitive and Individualistic Learning*. Needham Heights, MA: Allyn and Bacon.

Kadushin, C. *Some Basic Network Concepts and Propositions (Draft)* 2004. http://home.earthlink.net/~ckadushin/Texts/Basic%20Network%20Concepts.pdf.

Kapferer, B. 1972. *Strategy and Transaction in an African Factory*. Manchester: Manchester Univ. Press.

Knoke, D., and J. Kuklinski. 1982. *Network Analysis*. Beverly Hills, CA: Sage.

Krackhardt, D. 1987. Cognitive Social Structures. *Social Networks* 9:104-34.

Martinez, A., Y. Dimitriadis, B. Rubia, E. Gomez, L. Garrachon, and J. A. Marcos. 2002. Studying Social Aspects of Computer-Supported Collaboration with a Mixed Evaluation Approach. In *Proceedings of Computer Support for Collaborative Learning (CSCL 2002) Conference*, edited by G. Stahl, 631-32. Mahwah, NJ: Lawrence Erlbaum.

Mayadas, A. F. 2000. *Testimony of A. Frank Mayadas, Program Director, Alfred P. Sloan Foundation, before the Web-Based Education Committee*. http://www.hpcnet.org/cgi-bin/global/a_bus_card.cgi?SiteID=179526.

Middendorf, M., E. Ziv, C. Adams, J. Hom, R. Koytcheff, C. Levovitz, G. Woods, L. Chen, and C. Wiggins. 2004. Discriminative Topological Features Reveal Biological Network Mechanisms. *BMC Bioinformatics* 5:181.

Milo, R., S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298:824–27.

Molloy, M., and B. Reed. 1995. A Critical Point for Random Graphs with a Given Degree Sequence. *Random Struct. Algorithms* 6:161.

Monge, P. R., and N. S. Contractor. 2003. *Theories of Communication Networks*. Oxford, UK: Oxford University Press.

NetMiner CYRAM Co. Ltd. 2004. NetMiner 2.5. http://www.netminer.com/NetMiner/home_01.jsp.

Newman, M. E. J. 2003. The Structure and Function of Complex Networks. *SIAM Review* 45 (2):167-256.

Newman, M. E. J., A. L. Barabasi, and D. J. Watts. 2003. *The Structure and Dynamics of Networks*. Princeton, NJ: Princeton University Press.

Newman, M. E. J., S. H. Strogatz, and D. J. Watts. 2001. Random Graphs with Arbitrary Degree Distributions and Their Applications. *Phys. Rev. E* 64:026118.
Oshima, J., C. Bereiter, and M. Scardamalia. 1995. Information-Access Characteristics for High Conceptual Progress in a Computer Networked Learning Environment. In *Support for Collaborative Learning '95*, edited by J. L. Schnase and E. L. Cunnius, 259-67. Bloomington, IN: Lawrence Erlbaum Associates.
Rafaeli, S. 1988. From New Media to Communication. In *Sage Annual Review of Communication Research: Advancing Communication Science*, 110-34. Beverly Hills, CA: Sage.
Rafaeli, S., and R. J. LaRose. 1993. Electronic Bulletin Boards and "Public Goods" Explanations of Collaborative Mass Media. *Communication Research* 20:277-97.
Rafaeli, S., F. Sudweeks, E. Mabry, and J. Konstan. 1998. ProjectH: A Collaborative Qualitative Study of Computer-Mediated Communication. In *Network and Netplay: Virtual Groups on the Internet*, edited by F. Sudweeks, M. L. McLaughlin and S. Rafaeli, 265-82. Menlo Park, CA: AAAI/MIT Press.
Rao, A. R., and S. Bandyopadhyay. 1987. Measures of Reciprocity in a Social Network. *Sankhy 49 Ser. A.*: 141-48.
Reffay, C., and T. Chanier. 2002. Social Network Analysis Used for Modeling Collaboration in Distance Learning Groups. In *Lecture Notes in Computer Science (LNCS)*, edited by S. A. Cerri, G. Guarderes and F. Paraguaco, 31-40.
Richardson, J. C., and K. Swan. 2003. Examining Social Presence in Online Courses in Relation to Students' Perceived Learning and Satisfaction. *Journal of Asynchronous Learning Networks* 7 (1):68-88. http://www.sloan-c.org/publications/jaln/v7n1/pdf/v7n1_richardson.pdf.
Seabright, P. 2004. *The Company of Strangers: A Natural History of Economic Life*. Princeton, NJ: Princeton University Press.
Skvoretz, J. 1990. Biased Net Theory: Approximations, Simulations, and Observations. *Social Networks* 12:217-38.
Skvoretz, J. 2002. Relations, Species, and Network Structure. *Journal of Social Structure* 3 (3).
Snijders, T. A. B. 1991. Enumeration and Simulation Methods for 0–1 Matrices with Given Marginals. *Psychometrika* 56 (3):397-417.
Walther, J. B. 1994. Anticipated Ongoing Interaction Versus Channel Effects on Relational Communication in Computer Mediated Interaction. *Human Communication Research* 20:497-526.
Wang, Y., and D. R. Fesenmaier. 2003. Understanding the Motivation of Contribution in Online Communities: An Empirical Investigation of an Online Travel Community. *Electronic Markets* 13 (1):33-45.
Wasserman, S., and K. Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.
Wasserman, S., and P. E. Pattison. 1996. Logit Models and Logistic Regression for Social Networks. Part I. An Introduction to Markov Graphs and p*. *Psychometrika* 60:401-26.
Wegerif, R. 1998. The Social Dimension of Asynchronous Learning Networks. *Journal of Asynchronous Learning Networks* 2 (1):34-49. http://www.aln.org/alnweb/journal/vol2_issue1/wegerif.htm.
Wellman, B., and M. Gulia. 1999. Net Surfers Don't Ride Alone: Virtual Communities as Communities. In *Communities and Cyberspace*, edited by P. Kollock and M. P. Smith. New York, NY: Routledge.

The Life Cycle of Collaborative Partnerships: evolution of structure and roles in industry-university research networks.

Robert T. Trotter, II
Northern Arizona University, Department of Anthropology

Elizabeth K. Briody
General Motors Research and Development Center, Vehicle Development Research Laboratory

Gülcin H. Sengir
General Motors Research and Development Center, Manufacturing Systems Laboratory

Tracy L. Meerwarth
General Motors Research and Development Center, Vehicle Development Research Laboratory

ABSTRACT

Global corporations have initiated collaborative partnerships with university research institutions, private-sector firms, and other strategic partners at an increasingly rapid pace over the last decade. These partnerships create collaborative networks that leverage knowledge acquisition and technology transfer necessary to keep corporations and universities at the cutting edge of competition. Consequently, corporations have a competitive need to be able to predict the ideal structure, dynamics, and life cycles of these partnerships in order to effectively initiate, maintain, repair, and exit them in a way that retains the potential for future collaboration for both sides of the partnership. This paper provides an empirically validated model of the evolutionary structures and role relationships found in successful collaborative partnerships. The research combined ethnographic methods with qualitative and quantitative social network paradigms to identify the key structural frameworks and role configurations critical to the health of partnerships over their typical life cycle. The results include a description of the structures and the key player dynamics of these partnerships through six life cycle stages (approach, initiation, start-up, growth, maturity, and transition).

Corresponding Author Information: Robert T. Trotter, II. Arizona Regent's Professor, Department of Anthropology, Northern Arizona University, Flagstaff, AZ 86011. Email: Robert.Trotter@nau.edu, phone: 928-523-4521 (cel. 928-380-8694): web site: http://jan.ucc.nau.edu/~rtt/

We would like to acknowledge the warm welcome we received when we made site visits to the four CRL universities, and the assistance they provided throughout the study. We appreciate the time that the GM and CRL participants spent with us in discussions and for their responses on the social-network survey.

INTRODUCTION

Global competition is accelerating the trend for corporations to leverage university knowledge and expertise through formal collaborative partnerships (Barringer and Harrison 2000, Neill et al. 2001). The overall goal of these partnerships is to spur diffusion of innovation (Sussman et al. 2006, Valente and Rogers 1995) and keep up with rapidly changing research needs (Lewis 2000). In the late 1990s, General Motors Research and Development Center (GM R&D) initiated a Collaborative Research Laboratories (CRL) strategy as a strategic initiative. Previously, connections between GM R&D and universities were based largely on pre-existing dyadic relationships between researchers or R&D contracts with specific professors. In 2002, GM R&D management requested an examination of the structure and functioning of their successful collaborative research partnerships to identify ways to maintain and improve their effectiveness.

While the industrial literature has been primarily directed towards general organizational evolution (Laszlo, 2001, Learned, 1992) or focused on inter-organizational theory and practices (Anderson et al.1994, Prescott et al. 1998), the need to explore the overall evolutionary processes of collaborative partnerships has been identified, but only moderately addressed (Ring and Van de Ven 1994, Doz 1996). Consequently, there are a few social network studies (Borgatti and Foster 2003) that explore the evolution of partnerships (Stuart 1998, Ahuja 2000), the durability of networks (Kogut and Walker 2001), longitudinal analysis of alliance formation (Gulati 1995), transitional networks (Madhavan et al. 1998), and the concept of social capital in the formation of industry networks (Walker et al. 1997). However, one of the gaps in this literature was the lack of a description of the structural and role based changes that might predictably occur over the life of a partnership.

Our initial CRL data analysis produced a cultural model of collaboration (Sengir et al. 2004, Trotter et al. 2004) that highlighted key patterns in the relationship dynamics of partnerships. The original systems dynamics model of these partnerships was focused on relationship conditions (trust, cooperation, conflict, communication, joint work, etc.) and was designed as a diagnostic tool for industry-university collaborations (Sengir et al. 2004), utilizing a life cycle baseline data set. This article provides a substantial enhancement of the original model by elaborating the key structural (network) and role functions imbedded in the original model. This article focuses on describing the stage-based evolutionary (life cycle) conditions found in successful collaborative partnerships. The hypothetical and empirical data presented in this article can be used to form a “best practices” model for this type of partnership.

METHODS

We employed three synergistic methodologies: 1) ethnographic studies at GM R&D and at four CRL sites, using standard applied ethnographic methods (Trotter and Schensul 1998); 2) a social network survey that allowed us to investigate partnership structures, dynamics, and roles in the target partnerships; and 3) qualitative reliability and validity checks of our findings through formal validation sessions (see Kirk and Miller 1986). Each approach followed a comparative-empirical analysis strategy focusing on themes and patterns (Bernard 1998, Schensul and

LeCompte 1999) informed by prior ethnographic research on partnerships (Meerwarth, Briody, and Kulkarni 2005), and informed by general network analysis theory (Wasserman and Faust 1994, Wellman and Berkowitz 1997), with an emphasis on the qualitative aspects of network relationships.

Ethnographic Research Methods

Ethnographic interview, focus group, observational and documentary data were collected at four collaborative labs (Zeta, Gamma, Delta, and Alpha Universities -- pseudonyms following standard ethnographic confidentiality conventions) and at GM R&D. The primary ethnographic methods included in-depth semi-structured and *in situ* interviews based on iteratively developed interview and focus group guides; direct observation of collaborative laboratories and the accompanying interactions between partners; participant observation of key processes; culture-in-context observations that identified the normative behavior at the various collaborative sites and venues; and focused qualitative data collection (free listings, cultural model interviews) on the meaning of collaboration, roles and role definitions and information on the formal and informal structures of the collaborative laboratories. This approach allowed us to describe the context as well as the basic cultural viewpoints on collaboration and social networks within and between the partner organizations. It also allowed us to develop and refine the key variables that we included in a social network survey of the partnerships.

We conducted in-depth ethnographic interviews with 65 individuals, 38 from GM and 27 from the partnering institutions. Ethnographic informants were selected using a nominated expert sampling process. The core research and administrative personnel at GM and the CRL (CRL Director, GM champion, GM and CRL thrust area leaders, department heads and chairs, etc.) were identified and interviewed (expert saturation sample). This core expert group then nominated additional individuals who were qualitatively representative of the whole "experience and expertise" configuration of the CRL, including graduate students, technicians, post-doc students, faculty, ancillary GM personnel, and administrative assistants. Our interview questions focused on the nature of the participants' past and current relationships with their counterparts, perceived success factors for and obstacles confronting the partnership, institutional/organizational cultures of the partners, individual roles that were important to the development and maintenance of the CRL, and expectations about the future of the partnership. Eight focus groups (average 8 persons each) explored partnership goals and expectations, the participants' current assessment of the partnership, recipes for an ideal partnership, and ideas for strengthening these long-term relationships. A set of 6 field observation studies provided data on interactions, key collaborative processes, and meetings both at GM and at the partnership institutions. Finally, CRL documents provided insight into the formally-stated goals and activities of these partnerships.

GM had established four CRLs at U.S. universities by 2002 when this research was initiated. The first Lab was established at Alpha University in 1998. This lab was nearing the end of its first partnership cycle and was exploring options for renewal. It provided us with baseline information about all of the key stages that CRLs experience and the transitions that are likely in the later part of the partnership cycle. In 2000, GM established the second CRL at Gamma University, which allowed us to investigate both the early stages and some middle stages of cooperation and the transitions faced during those times. Delta University became the third collaboration early in 2001. The fourth CRL at Zeta University also commenced in 2001, several months after the GM-Delta University. Both of these partnerships provided extensive data on the selection and recruitment stages of the partnership life cycle, and solid information on the start up stage. The overall data set provided details on the ways successful partnerships are initiated, gain momentum, and go through end of cycle transformations. It also provided information on the changes in individual roles, numbers of participants and types of participation that are necessary for a healthy partnership life cycle (Sengir et al. 2004, Trotter et al. 2004).

Social Network Survey

We administered two email-based social network surveys to both the GM and CRL participants. The surveys were sent to every individual who was identified by either GM or the CRL as being involved in any role or activity in the partnership. The first survey was sent to 176 participants in the original four CRLs and followed general network data collection guidelines (cf. Wasserman and Faust 1994). The instrument included demographic questions (name, position, and experience with collaborative relationships) followed by a general "name generator" matrix requesting each respondent to identify the complete list of individuals that they were in contact with as part of the CRL. For each individual named, alters were ranked on perceived levels of communication, joint work, trust, conflict, and cooperation.

The survey response rate of 62.5 percent (of the 176 surveys delivered) was methodologically acceptable, based on an expected response rate of approximately 35 percent (Stork and Richards 1992). The second survey was conducted 18 months later, as a "time two" validation test for the general model. At that time, there were a total of 8 CRLs in operation and 270 surveys were sent out, with a response rate of 68.1 percent. There was only one active refusal to participate in either of the surveys. The only observable difference between the response and non-response groups was a trend towards a lower response rate among the more peripherally involved graduate students and technicians compared with faculty, post docs, GM researchers, and administrators from both sides of the partnership. The trend does not appear significant and does not appear to have had a substantial impact on either the structural or the key player data presented below.

The analytical techniques applied to the survey data included free listing, egocentric, sociometric, and network visualization analysis. The free listing analysis provided information on changes in the size, content, and configuration of named relationships (Weller and Romney 1997). The egocentric analysis allowed us to compare individual role types and institutional groups (GM participants, University participants, etc.) for a range of variables including communication, trust, conflict, and work importance, among others, following Borgatti et al. (2002a, 2002b), and Cross et al. (2002) The sociometric data allowed us to construct network structures for visualization analysis, sociometric comparisons of the networks at various life cycle stages, as well as individual roles, subgroups, and measures of association and communication. We utilized one ethnographic (ANTHROPAC 4.98: Borgatti 1996) and four network programs (UCINET 6: Borgatti et al. 2003, Key Player: Borgatti 2002b, NETDRAW: Borgatti 2002a, Mage: Richardson 2001 for 3-D visualization). The network visualization process allowed us to compare the structures of the various CRL networks within the framework of the life cycle stages established in the ethnographic data. The role analyses combined ethnographic and network data using both Key Player software (Borgatti 2002d) and sociometric and visualization analysis of the survey data (Borgatti 2002c). We also utilized the visual analysis characteristics of NETDRAW (Borgatti 2002a) to identify key positions and subgroups in the CRL networks at various stages in their life cycle.

Validation and Triangulation Process

We conducted 10 independent validation sessions attended by 145 study participants. We presented findings and gathered input on data validity and any potential gaps in data collection. These sessions were designed to qualitatively test the soundness of our analyses, to integrate new insights into our work, and to collect additional data (Kirk and Miller 1986, Bernard 1998). This validation process is a hallmark of strong ethnographic projects and provides the qualitative equivalent of the reliability and validity testing conducted in any well designed quantitative project.

RESULTS

All CRL partnerships undergo a selection and approval process, followed by a start up period, a growth period, a mature stage, and an end of cycle transition stage. Table 1 briefly summarizes the key characteristics of the stages, as well as some of the predominant characteristics of their network structure and key player (role) dynamics.

Table 1: Summary of Defining Characteristics: Structural Elements and Role Dynamics of Successful Collaborative Research Partnerships

Partnership Stage	Defining Characteristics	Structural Characteristics of Collaborative Networks	Role (Key Player) Characteristics of Collaborative Networks
Approach	Informal exploration of mutual interests; formal requests for statement of interest	Mostly dyads and triads in informal discussions	Management and administrative roles predominate
Initiation	Formal negotiation of goals, processes, intellectual property issues	Small densely connected work groups	A mix of managerial, technical, and support roles representing both sides of the partnership.
Start Up	Creation of core partnership membership; establishment of key relationships	Relatively small core-periphery structured network; high density, strong ties predominate	Key players create and maintain a core group that will be relatively stable throughout the partnership life cycle
Growth	Consolidation of relationships; initiation and elaboration of collaborative work processes, initiation of joint work productivity	Growing core-periphery structure group with core maintaining goals and direction of partnership and peripheries increasingly focused on specific work tasks	Increased differentiation and growth in key players. Key players in core primarily serving integrative functions, key players in periphery structures acting as bridges, catalysts, etc.
Mature	Fulfillment of common goals; maintenance of core values, direction; full focus on productivity	Core and periphery structure elaborated into distinct subgroups (subgraphs) primarily focused on joint work; core focused on integrative roles	Key player roles have increased emphasis on problem solving and adjudication, as well as integrative roles. Key player roles in subgroups are elaborated.
Transition	Assessment Period focused on quality and outcomes of partnership (goals met and unmet); relationship dynamics reviewed; risks to continuation, modification, and termination assessed	Mature structure is maintained up to actual transition or is modified over a relatively short period leading up to transition. Four end-game structures are possible: minimal change, added or subtracted thrust areas, split and increase number of CRLs, or close down.	Tension and ambiguity lead to threat to quality of relationships; Key roles modified from maintenance to transition roles (emphasis on damage control, problem resolution, or revitalization); Temporary or permanent reduction or suspension of technical roles.

The following two sections combine the basic stage descriptions from our interview and observational data with the structural and role data from the social network surveys to describe the conditions that apply to successful collaborative partnerships through their life cycles. The first section emphasizes the evolution of basic network structures that are found in successful partnerships. The second section provides information on the role and position data that helps define each stage. The result is a complex and detailed, but utilitarian model for the primary network elements of a collaborative partnership.

Structural Characteristics of Successful Collaborative Networks

Each of the life cycle stages can be identified and differentiated from the others on the basis of their network characteristics, including size and growth, differences in their core-periphery structures, and differences in positions necessary to the functioning of the partnerships at each stage.

Approach Stage: The earliest interactions between GM and potential CRL partners constitute the Approach Stage. This stage begins with the identification of a specific GM R&D research need. R&D personnel then nominate prominent universities and programs that are leaders in the area of interest. Once a potential field of candidate schools is identified, each is contacted, usually through existing connections between GM R&D personnel and the respective universities. There is a brief period of informal interaction between key players in GM R&D and the various university key players to explore the initial level of mutual interest. The participant group is then narrowed to include one or two programs that are requested to enter into formal discussions. The primary network connections for this stage are weak ties (often based on interactions at scientific meetings) or strong dyads (school ties of both the relationship sort, and the sartorial variety).

Courtship Stage: The Courtship Stage encompasses formal and informal negotiations about the specific goals and structures of the partnership. The Courtship Stage begins with general negotiations and ends with a joint identification of thrust areas and a formal Agreement. A small number of individuals explore common ground (potential joint work) and negotiate key institutional concerns such as intellectual property rights, resources, and commitment of personnel. Key players begin to emerge on each side of the partnership. One key player described how GM and Delta University worked out many of the details of their new formal relationship:

> *It took us almost five months to develop the contract…Those five months were very important…I couldn't be happier that we spent those five months. They defined what the deliverables were… how we are [were] going to approve different projects and propose different projects, what were the intellectual property issues that we need[ed] to deal with. Who does what, basically and also figuring out what objectives we will be following? That time and planning was very, very helpful to us.*

Initiation Stage: The first deliberately constructed networks begin at the initiation stage. Our qualitative interview data allows us to represent the Initiation Stage visually as a ladder (Figure 1). This is a typical business organizational structure that involves individuals in administrative roles (hierarchical organization -- top of ladder) and research roles (horizontal organization--

lower rungs of ladder). The ladder shape, rather than the dendrogram shape of a standard organizational chart, results when the partners deliberately include one or more individuals from each key level of both organizations and formally pair them with their counterparts at the same organizational level. This hypothetical structure was deliberately created for each CRL, with the expressed hope that additional vertical and horizontal cross-connections would rapidly follow. Individuals from the two partnering organizations are represented by red or blue nodes, respectively.

Figure 1: Hierarchical Initiation Phase Network "Ladder Configuration"
[Hypothetical Construction]

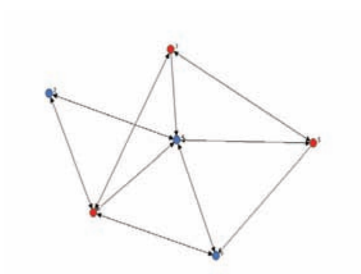


Figure 2: Initiation Phase Configuration with two Pre-Existing Relationships
[Empirical Qualitative Data]

Figure 2 illustrates a typical reconfiguration of the original ladder hierarchy that reduces the rigidity of the initial structure. The successful partnerships had at least one or two people who were previously connected through their work, but were not at the same hierarchical level in their respective organizations. When those individuals are included in the start up process for the partnership, their ties cross connect between levels and across organizations, changing the structural configuration of the partnership. Figure 2 depicts an Initiation Stage partnership in which there are two pre-existing relationships among the individuals initiating Start-Up. This non-ladder configuration is more effective for rapidly developing the necessary partnering relationships than the hierarchical structure. Communication flow and decision making can be faster and more consensual, avoiding "red tape" and other bureaucratic impediments at a critical stage of partnership development. The overall demands of collaboration require that people talk and work with one another up, down, and across both organizational hierarchies.

Start-Up Stage. The Start-Up Stage emphasizes relationship dynamics and the interactions that hold the overall collaboration on course, including communication, the development of trust, and overall positive reciprocity between individuals (reduction of conflict, initiation of cooperation). One participant stated:

> *The people who end up working together need to understand and appreciate each other. They need mutual respect and this is the major element of success for us.*

Another participant described the relationship-development process in this way:

> *In my area it has taken these two years to establish a real good collaborative collegial relationship. It takes regularly attending [working meetings] to get out of it what we should be getting out of it…So, we drive down every few weeks [to Gamma] and we go to the quarterly reviews.*

Figure 3 is a hypothetical network structure comprised of 15 individuals (modal number for start up groups) constructed from ethnographic descriptions of the relationships our informants felt should ideally exist at Start-Up. Figures 4 and 5 are empirical network representations of two CRL networks surveyed at the Start-Up stage. The globular structure (core-periphery structure in social network terminology) of all three networks focuses the efforts of one or two key players who are connected through communication and interactions with everyone else in the network. This structure becomes the enduring network glue that holds the collaboration together throughout the partnership life cycle.

Figure 3: Start Up Structure [Hypothetical Construction]

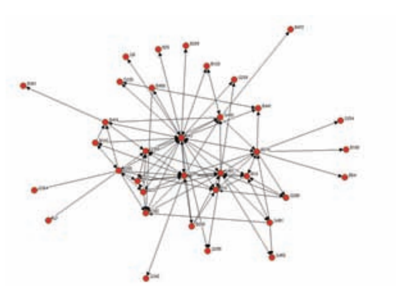


Figure 4: Start-Up Structure at 1 Year (GM-Zeta-empirical)

Figure 5: Start-Up Structure at 1.5 Years (GM-Delta-empirical)

The core-periphery structure of all start up partnerships has a common condition visible in figures 4 and 5. The core contains a group of people who are densely connected across both sides of the partnership. The periphery contains some individuals with a single connection or tie to one of the core members. This gives the structure a "prickly" look, from a qualitative perspective. The qualitative data indicated that these single connection individuals are usually either graduate students (on the university side) who are tied to the overall partnership by a single faculty member, or they are technicians or GM researchers who have a single tie (due to their specific expertise) to only one of the GM participants. One visible difference between the partnership stages occurs when these single tie individuals develop connections to each other and to other core individuals. This process elaboration of the number and complexity of ties at the periphery is a key condition that defines the difference between the Start Up and Growth stages for the network data.

<u>Growth Stage</u> The Growth Stage begins when stable core relationships allow joint work processes to emerge as distinct subgroups within the overall partnership. During the Growth Stage the partnership emphasis is on increasing productivity, in addition to maintaining positive relationships. At this stage "thrust areas" (formally established technical areas for specific joint research collaboration) emerge, increase in size, and form visible sub-groups (subgraphs). They protect their localized dynamics by establishing key player gatekeepers who keep the demands from the overall partnership relationship reasonable, while increasing the productivity in the subgroup. One participant commented:

> *We've established a closer interaction. This is due to the maturity of the program. Now, we*

> *are working on stuff. It would have been less helpful to have more [technical] interactions earlier.*

Figures 6 and 7 illustrate the hypothetical and empirically derived structures for the Growth Stage. In Figure 6, two thrust areas are beginning to develop as new participants are added to the core (represented by red nodes). These thrust areas are represented by a cluster of yellow and blue nodes, respectively, at the peripheries. These thrust areas show increasing local density separate from the connectivity of the core. Figure 7 illustrates emerging thrust area structures at the "northwest," "southeast" and "northeast" quadrants of the Gamma network visual data.

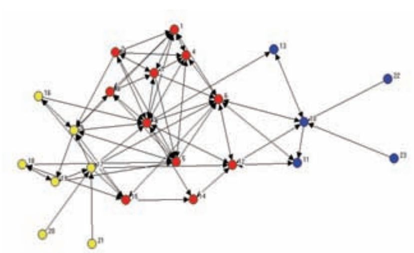


Figure 6: Growth Stage Structure with two Emerging Thrust Areas [Hypothetical Construction]

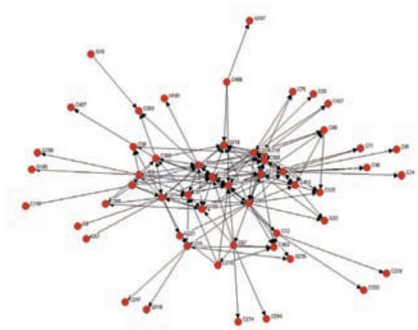


Figure 7: Growth Stage Structure at 3 Years with three Thrust Areas (GM-Gamma-empirical)

This structure is technically also a core-periphery network structure similar to the start up stage structure (or a continuation and elaboration of it). It qualitatively differs from the start up stage as the periphery visually shows the growth of the whole network and the elaboration of localized subcomponents.

<u>Mature Stage</u>. The mature stage is the highest productivity stage and has the most complex structure. It allows productivity to be increased or maintained, while balancing the need for overall integration of the partnership through common vision and goals. There is a continuing effort by core key players to keep the collaboration on track (integration), complemented by focused productivity to meet joint work goals. One participant stated:

> *Above the thrust areas, there is the integration function. If we do something in one area, we want to know how this will affect other areas and how it will affect how GM does business.*

The structural result of these complementary processes (integration and productivity) is represented in Figure 8. Visually the overall structure looks like a boat propeller, fitting nicely with the metaphor behind thrust areas and positive forward directions for the partnership. The integrative core is displayed in yellow (the hub of the propeller), and three mature and productive thrust areas are illustrated by red, blue, and green nodes respectively. While reality is somewhat messier than the ideal, the visual presentation of the empirical data (Figure 9) reveals the core and periphery structure is sufficiently close to the ideal to state that the hypothetical structure has appeared in a real world situation. Figure 9 also illustrates that some individuals at the peripheries are connected to only one other individual in the fan structure. This is the same condition demonstrated in the start up and growth phase. This data suggests a growth pattern that continues to occur throughout partnership. The overall structure suggests that innovation for these partnerships typically moves from the peripheries to the core, with the core controlling technological transfer into the broader institutions that support the partnerships.

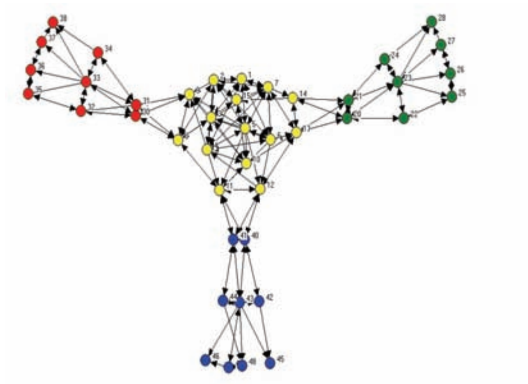


Figure 8: Mature Structure [Hypothetical Construction]

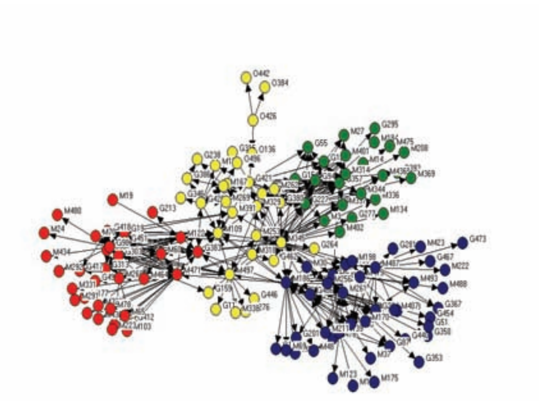


Figure 9: Mature Structure at 4.5 Years (GM-Alpha-empirical)

Transition Stage. The Transition Stage represents the end of the formal CRL life cycle. The processes that govern both the unilateral and bilateral decisions about the partnership come into prominence at this stage and potentially threaten the relationships and the networks that have been created. One or both partners become concerned over transition decisions. One CRL participant stated,

> *We are in the fourth year of the partnership and starting the fifth. The funding runs out in 2002. We have built a mechanism and an infrastructure for this work. It would be good to know ahead of time if we'll be renewed by GM. We've got students lined up that need the support.*

Transition issues often refocus the emphasis of the partnerships away from joint work and back to the core structure at the center of the partnership. Relationships become more ambiguous and prone to misinterpretation or negative interpretation. Conflict can arise based on both rumored and actual changes, and "whole group" communication becomes important. Four transition options were identified in the ethnographic data. The CRL can continue in its original form, as was the case for the GM-Gamma CRL when it was renewed at the end of three years. The existing structure and key player roles continued largely without interruption. A second option is to modify the CRL by adding or eliminating thrust areas, producing a reconfigured structure and key player configuration. Typically one or more thrust areas are disbanded; alternately, one or more thrust areas may be added. The GM-Delta CRL represents this option where three out of four thrust areas were abandoned by mutual consent and the fourth was continued. A third option is to split a mature CRL into two or more independent CRLs. This option occurred with the GM-Alpha CRL. One of the original three thrust areas was dissolved, and the two remaining were allowed to separate and form two new CRLs with multiple thrust areas. The fourth option is for the CRL to be terminated. If the termination process is conducted appropriately, the formal structure of the CRL will disappear, but many of the key dyadic relationships persist, and the overall partnership experience is judged to have been positive and productive.

The following section presents the data on the changing roles and positions that simultaneously occur throughout the CRL life cycle, in conjunction with the structural changes described above.

The Key Player Mix: Changing Roles and the CRL Life Cycle:

The GM and CRL participants provided substantial qualitative and sociometric data on key player roles within the overall CRL life stage model. Both the interview and observational data emphasizes the importance of these individuals, without whom the partnerships would have foundered. One participant stated,

> *You absolutely have to have people who provide leadership. Leaders are individuals who are aware of what's going on in the program and who are providing leadership to the program, but they are also providing monitoring. They are very, very critical to the success of the program because they are willing to identify where people are making contributions, and identify and reward those contributions. But they are willing to identify people who are not making contributions [also].*

Following Borgatti (2002b) we identified leaders through a "key player" analysis, informed by our qualitative data on roles and positions. We operationally defined key players as individuals who take on critical roles in the formation and maintenance of CRL networks. We compared their sociometric positions with the qualitative data we had available on both the individuals and their roles.

The CRLs have at least one, and more often two or three individuals whose primary role is to keep communication lines open, solve problems, and help solid relationships develop or be maintained throughout the partnership life cycle. One participant commented:

> *What I've learned is that it's essential to have a committed person at Alpha and at GM. The partnership is going to survive or fall on the personal interactions between these two people.*

This role continues throughout the CRL life cycle, supported by the accretion of additional key players who stabilize and solidify the functional aspects of the network structure. Additionally, it is possible for key players to begin in one role, and as the CRL changes, for them to adapt or change their roles and remain key players throughout the life of the partnership. Others may not be successful in changing roles, and may need to be removed to improve the health of the partnership. One participant stated:

> *Maybe they (individuals not making contributions) were originally, but their contributions faded through time and they should move them off of projects and keep the energy and the productivity of the project up.*

Three types of key player analyses (reach, fragmentation, and cut points) are very useful for understanding the organizational-role aspects of partnership life cycles. The evolutionary aspects of the integrative role are described below in our "reach" analysis section. CRLs also have key players whose function is to stimulate and direct work activities in subgroups within the network. If these individuals are removed, there is an immediate need to "repair" the network to keep it meeting work related goals. We found that a "fragmentation" analysis of the CRLs was very useful in identifying key players whose replacement was very high priority if they left the network for some reason. Finally, we found it very useful to use the concept of "cut points" to identify the bridges to subsegments of the CRL networks. This allowed us to potentially match the organization roles and responsibilities of key players to the empirical data on their position in the CRL network structures, to see if any changes were needed. It also identified parts of the overall structure that were "natural" cut points during the transition phase of the CRL.

Reach: Ability to Easily Communicate with or Influence All CRL Participants.

We conducted a "reach" analysis (proportion of the network each individual is in contact with) to identify the key players who establish or maintain the maximum connection with alters in a network. Reach is one way of indirectly estimating the relative amount of time and effort that are necessary for getting accurate information to everyone in a network, as well as estimating the minimum number of people who need to adopt this role for different sizes of partnerships. One CRL participant described an individual filling the "reach" role of a key player:

> *{He} does an excellent job of keeping us informed, and involved, and his faculty involved. During the [joint meetings], he does an excellent job of presenting to us, bringing in others from outside his department, and that has led to some relationships.*

Table 2 identifies the extent of reach of one, two or three key players who have the maximum unique reach for their networks. Newer and smaller networks, such as GM-Zeta and GM-Delta, have single individuals, or at most pairs of individuals, who can contact everyone directly, or through only one intermediary. More mature networks, such as GM-Gamma and GM-Alpha, typically must utilize three or more people to make all of the linkages work.

Table 2: Stage Based Analysis of Reach: The Impact of Time and Network Size on Reach in Successful Collaborative Partnerships

CRL	Partnership Stage	Network Reach, One Key Player	Network Reach, Two Key Players	Network Reach, Three Key Players
GM-Zeta	Start-Up	**100** (U-1)	**100** (U-1, GM-1)	**100** (U-1, GM-1, GM-2)
GM-Delta	Late Start-Up	**89.6** (U-1)	**100** (GM-1, U-1)	**100** (GM-1, U-1, GM-2)
GM-Gamma	Growth	**92.7** (GM-1)	**98.2** (GM-1, U-1)	**100** (U-1, GM-1, GM-2)
GM-Alpha	Mature	**89.6** (U-1)	**97.8** (U-1, U-2)	**100** (U-1, GM-1, U-2)

The "U" and "GM" designations indicate which side of the collaboration (U for university, GM for GM) that the persons represent. The numbers (1, 2) represent the order in which the person appeared in the reach data.

The key player with the greatest amount of reach is typically from the university rather than GM side of the partnership. When reach is calculated for two key players, both university and GM key players emerge, with the exception of Alpha. At the most complex stage, at least one GM participant is required to achieve 100 percent reach. This finding emphasized the need for the partnerships to be truly collaborative, rather than to follow a market model of buying knowledge, since any lack of appropriate attention to the key players on the university side are very likely to result in reduced effectiveness and productivity for the partnership and a loss of knowledge and technology transfer for GM. Our analysis also demonstrated that there is considerable redundancy (i.e., overlapping reach) in the networks. Commonly, two or more individuals share very similar sets of relationships even though one has slightly more reach. This redundancy helps protect the network against problems produced by the loss of key individuals. Reach analysis can be used to identify individuals who would be good role or position replacements for other individuals, everything else being equal, because their reach "footprint" is virtually identical to the person being lost.

<u>Fragmentation</u>

One threat to CRL health and productivity comes from the loss of key players and the subsequent fragmentation of the collaboration. This threat is exacerbated by the natural development of clique-like subgroups in any longer term network. One participant commented,

> *There's a very natural tendency for two institutions to set up a collaborative project and then have that collaboration naturally fragment or naturally segment.*

In early partnership stages, losing virtually any key player from the core structure translates into serious fragmentation or even destruction of the partnership. In later stages, individual loss is less damaging, although the loss of multiple key players is still problematic. Following Borgatti (2002b), fragmentation is defined as the removal of a key player from a network when their removal means that individuals or other subunits in the network are no longer connected to the network as a whole. Stage-based fragmentation is illustrated in Table 3, which shows the levels of fragmentation caused by the removal of the highest impact one, two or three persons respectively in each CRL network.

Table 3: Stage Based Impact of Fragmentation in CRLs: Removal of Highest Impact Key Players

CRL	Partnership Stage	Fragmentation* One Key Player Removed	Fragmentation Two Key Players Removed	Fragmentation Three Key Players Removed
GM-Zeta	Start-Up	**0.21** (U-1)	**0.40** (U-1, U-2)	**0.56** (U –1, U-2, U-3)
GM-Delta	Late Start-Up	**0.36** (U-1)	**0.48** (U-1, GM-1)	**0.58** (U-1, GM-1, GM-2)
GM-Gamma	Growth	**0.11** (U-1)	**0.21** (U-1, GM-1)	**0.23** (U-1, GM-1, U-2)
GM-Alpha	Mature	**0.14** (U-1)	**0.24** (U-1, U-2)	**0.33** (U-1, U-2, GM-1)

*A fragmentation value towards 1 indicates the loss of the particular individual has created many small clusters of people such that the network is highly fragmented; a value toward 0 means that most nodes are still connected within the network (cf. Borgatti 2002d).

The survey data is consistent with the qualitative data. The new CRLs are more dependent on one or two central individuals than the older established CRLs. The impact of removing a single key player is higher in GM-Zeta and GM-Delta), than in GM-Gamma or GM-Alpha. The more established CRLs have more complex core-periphery structures that provide some protection against the "whole network" impact of fragmentation. Repairs to the network can proceed more rapidly in a more established CRL. On the other hand, the "fracture" points identified by the fragmentation data can also be used to identify individuals within the overall structure to target during the transition stage of the CRL life cycle, where special care must be taken to maintain a key relationship. This data is also consistent with the reach data, above. The highest-impact key player in any CRL, at any stage, is normally a university key player. This condition provides some leverage and influence for the university that is a counterweight to the fact that GM is providing the bulk of the resources that fund and support the partnership. As two or more key players are identified for any CRL, they tend to represent both the university side and the GM side of the partnership; both sides are critical to cohesion and success as the CRLs pass through the various stages of the partnership cycle.

Bridges and Cut Points

Some key players act as primarily as bridges to distinct subgroups in the CRL networks. These positions, sometimes called "cut points," link distinct segments or regions of the network. If they are removed a new bridge must be formed or contact will be truncated or lost with part of the network. Figures 10 and 11 visually identify cut points (red nodes) in a new and a mature CRL.

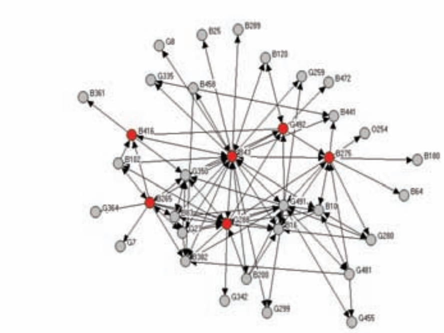


Figure 10: Cut Points and Bridges in a Start-Up Stage CRL (GM-Zeta)

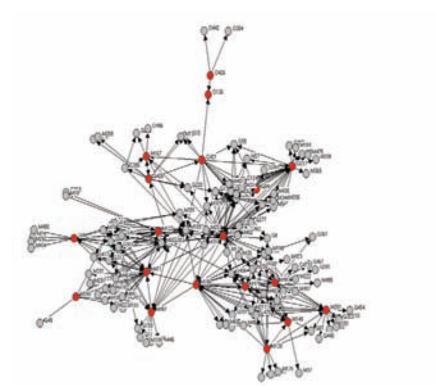


Figure 11: Cut Points and Bridges in a Mature Stage CRL (GM-Alpha)

The number of bridges needed in any CRL increases with both the size and the complexity of the CRL at each stage. The Start-Up Stage CRL contains six cut points that bridge its smaller segments and more homogeneous structure. The Mature Stage CRL contains about three times as many points (17) that bridge the more numerous sub-components embedded in the overall network. University cut points outnumber GM cut points throughout the partnership cycle, for each of the CRLs. During Start-Up, Zeta had four persons occupying cut points while GM had only two. During the Mature Stage, Alpha individuals occupied 10 cut points while GM personnel occupied seven. While this finding has to be considered preliminary, the consistent trend in our data suggests that there may be important differential contributions from the two sides of a collaborative partnership, depending on the nature of the participating organizations. This is an area that we intend to investigate further.

DISCUSSION

Our overall goal was to produce an empirically tested model illustrating how key social network elements of successful collaborative research laboratories change over time. The following conclusions and implications were both presented and validated in our ten formal validation sessions and are currently being used as "best practices" by the CRLs, since the practical use of the model is to describe the critical characteristics of collaborative partnerships that can be used to both replicate successful collaborations and to diagnose and address problems in failing partnerships.

CRLs grow in size and structural complexity over the course of the partnership even though the resource base for the partnerships remains unchanged. Recently established CRLs show a lower connectivity between pairs of individual participants, as indicated by the average distance between dyads, than do the more mature CRLs. Increasing connectivity in the early stages is a critical function of the core key players. This suggests that a significant start up period, to increase the strength of ties in the CRL, is necessary for success. Relatively informal communication methods and styles operating at the outset of the CRL (e.g., impromptu discussions, informal polling of opinions) are gradually replaced by more formal patterns, and the informal processes appear to be less effective as the CRL ages. Mature CRLs require more structured and pre-planned communication methods. The more mature CRLs have dense working subgroups, which maintain a sense of community, but their structure reduces the overall connectivity in the CRL as a whole.

Structural similarities and differences by partnership stage suggest that CRLs require continuity, growth, and role flexibility for critical human resources and task allocations as they age. A core of participants whose turnover is low helps to stabilize the CRLS throughout the partnership cycle. All of the CRLs, regardless of life cycle stage, have at least one, and more often two or three individuals whose primary role is to keep communication lines open, solve problems, and help solid relationships develop and be maintained. Without their efforts, CRL work would be much less successful because the coordination of CRL activities, resources, and deliverables, including oversight of the technical work, would be lacking. At the same time, elaborating the connections within and between the thrust areas and the core makes the CRL stronger and more productive, and ultimately is the structure that achieves the primary goals of the partnership.

Key player and role analysis indicate the actual structure of relationships in these partnerships is compatible with, but is not dominated by formal hierarchical organizational structures. This "reconfiguration" from the standard organizational chart is one of the strengths of the partnerships. Recently-created CRLs are more susceptible to damage if key players leave the network (e.g., due to retirement, job transfer, loss of interest) compared with more mature CRLs. Newer CRLs are largely dependent on one or two key players to hold the network together. By contrast, more established CRLs do not experience the same degree of fragmentation, based on single individual personnel changes. If a key player leaves an older CRL, the network is able to adjust more rapidly than a newer CRL.

There is a need in the more mature CRLs to both recognize and reward individuals who are changing roles, or taking on roles that are not as visible as they would be in the young CRLs. These differences can be used to change or target the way in which the CRLs are managed at different stages, and the way that problems are addressed. For example, individual key players in the newer CRLs have a higher degree of "reach" and the simpler structures of new CRLs make it relatively easy to contact and communicate with all CRL participants through informal means. More mature networks experience a lower degree of reach since they typically require a minimum of three people to ensure complete contact within the total network. The combined network reach of at least one key player from each side (GM and University) is necessary for complete "reach." This information can be used to determine the ways in which goals and accomplishments can be communicated to the CRLs, as well as ways in which emerging problems can be addressed through either formal or informal organizational interventions.

We feel that these details and elements of our elaborated model of successful networks will allow a direct application of ethnographic and network paradigms to the process of establishing, monitoring and maintaining existing and emergent collaborative partnerships for the future.

References

Ahuja, G. (2000). Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study. Administrative Science Quarterly, 45, 425-55.

Anderson, J.C.,Hakansson H., Johanson J. (1994). Dyadic Business Relationships within a Business Network Context. Journal of Marketing, 58, 1-15.

Barringer, B.R., and. Harrison J.S. (2000). Walking a Tightrope: Creating Value Through Interorganizational Relationships. Journal of Management, 26, 367-405.

Bernard, H.R., (1998). Handbook of Methods in Cultural Anthropology, Walnut Creek, CA: Altamira Press.

Borgatti, S.P. (1996) ANTHROPAC. Ver. 4.98 Boston, MA: Analytic Technologies..

Borgatti, S.P. (2002a). NETDRAW. Boston, MA: Analytic Technologies. .

Borgatti, S.P. (2002b). Key Player. Boston, MA: Analytic Technologies. .

Borgatti, S.P. (2002c). A Brief Guide to Using NetDraw. Boston, MA: Analytic Technologies.

Borgatti, S.P. (2002d). KeyPlayer 1.1 User's Guide. Boston, MA: Analytic Technologies.

Borgatti, S.P. and Foster, P. (2003). The network paradigm in organizational research: A review and typology. Journal of Management, 29, 991-1013.

Borgatti, S.P., Everett M.G., Freeman L.C. (2003). UCINET 6. Harvard, MA: Analytic Technologies.

Cross, R., Borgatti, S.P., Parker, A., (2002). Making Invisible Work Visible: Using Social Network Analysis to Support Strategic Collaboration. California Management Review, 44, 25-46.

Doz, Y. (1996). The Evolution of Cooperation in Strategic Alliances: Initial Conditions or Learning Processes? Strategic Management Journal, Summer Special Issue, 17, 55-84.

Gulati, R. (1995). Social Structure and Alliance Formation Patterns: A Longitudinal Analysis. Administrative Science Quarterly, 40, 619-52.

Kirk, J. and Miller M.L. (1986). Reliability and Validity in Qualitative Research. Beverly Hills, CA: Sage Publications.

Kogut, B. and Walker G., (2001). The Small World of Germany and the Durability of National Networks, American Sociological Review, 66, 317-335.

Laszlo, A. (2001). The Epistemological Foundations of Evolutionary Systems Design, Systems Research and Behavioral Science, 18, 307-321.

Learned, K.E. (1992). What Happened Before the Organization? A Model of Organization Formation. Entrepreneurship:Theory and Practice, 17, 39-49.

Lewis, S.R. (2000). Corporate Partnerships Define the New R&D. R & D, 42, 1-12.

Madhavan, R., Koka B., Prescott, E. John. (1998). Networks in Transition: How Industry Events (Re) Shape Interfirm Relationships. Strategic Management Journal, 19, 439-459.

Meerwarth, T.L., Briody E.K., Kulkarni, D.M.. (2005). The Discovery and Exploration of Partnership Rules: A Methodological Perspective. Human Organization Human Organization, 64, 286–302.

Neill, J.D., Pfeiffer G.M., Young-Ybarra, C.E.. (2001). Technology R&D Alliances and Firm Value. Journal of High Technology Management Research, 12, 227-238.

Prescott, J. E., Koka B., Madhavan, R.. (1998). Networks in Transition: How Industry Events (re)Shape Interfirm Relationships. Strategic Management Journal, 19, 439-460.

Richardson, D.C. (2001). MAGE. Duke NC: Little River Institute, Duke University.

Ring, P.S. and Van de Ven, A.H.. (1994). Developmental Processes of Cooperative Interorganizational Relationships. Academy of Management Review, 19, 90-118.

Schensul, J.J. and LeCompte M.D. (1999). Designing and Conducting Ethnographic Research. Vol. 1, Walnut Creek, CA: Altamira Press.

Sengir, G.H., Trotter, R.T II, Briody, E.K, Kulkarni, D.M, Catlin, L.B., Meerwarth, T.L.. (2004). Modeling Relationship Dynamics in GM's Research-Institution Partnerships. Journal of Manufacturing Technology Management, 15, 541-559.

Stork, D. and Richards, W.D., (1992). "Nonrespondents in Communication Network Studies: Problems and Possibilities. Group and Organization Management, 17, 193- 210.

Stuart, T.E. (19980. Network Positions and Propensities to Collaborate: An Investigation of Strategic Alliance Formation in a High-Technology Industry. Administrative Science Quarterly, 43, 668-98.

Sussman, S, Valente, T.W. Rohrbach, L.A., Skara S., Pentz, M.A. (2006). Translation in the Health Professions: Converting Science into Action. Evaluation & the Health Professions, 29, 7-32.

Trotter, R.T. II, and Schensul, J.J. (1998). Methods in Applied Anthropology. In Handbook of Methods in Cultural Anthropology (pp. 691-736), H. Russell Bernard, ed., Walnut Creek, CA: Altimira Press.

Trotter, R. T. II., Briody, E.K., Sengir, G.H, Meerwarth, T.L., Catlin, L.B. (2004). The Evolving Nature of GM R&D'S Collaborative Research Labs: Learning from Stages and Roles. Warren. MI. GM Research & Development Center. R&D Publication No. 9907. 15 October.

Valente, T.W., Rogers, E.M. (1995) The Origins and Development of the Diffusion of Innovations Paradigm as an Example of Scientific Growth. Science Communication, 16, 242-273.

Walker, G., Kogut, B., & Shan W. (1997) Social Capital, Structural Holes and the Formation of an Industry Network. Organization Science. 8, 109-125.

Wasserman, S. & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge: Cambridge University Press.

Weller, S. & Romney, A.K.. (1997) Systematic Data Collection. Menlo Park, CA: Sage Publications.

Wellman, B. and Berkowitz, S.D. eds. (1997) Social Structures: A Network Approach Greenwich, CT: JAI [Elsevier] Press

The Eurovision Song Contest as a 'Friendship' Network

Anthony Dekker

Australian Defence Science and Technology Organisation (DSTO)

Abstract: *This paper examines the votes cast in the 2005 Eurovision Song Contest. Adjusting votes for song quality, a friendship network with valued links is obtained. Statistical analysis shows that friendship between countries is largely determined by geographical proximity, with a visible five-bloc structure. However, large immigrant groups often swayed national ties by voting for their home country. Some countries, such as Switzerland, appear to play a significant bridging role, and the Eastern Mediterranean bloc appears to act as a bridge to the new Balkan countries. Analysis thus reveals an emerging Europe very different from previous network studies of this kind. The analysis techniques demonstrated here have more general applicability, and may be useful for analysing other types of friendship networks.*

INTRODUCTION

The Eurovision Song Contest has been held annually since 1956. Hosted by the European Broadcasting Union, and broadcast live on television across Europe (with delayed telecasts internationally), the Eurovision Song Contest seeks to find Europe's most popular song. Perhaps the most famous winner has been Abba, the Swedish entry in 1974, singing "Waterloo." On 21 May 2005, the 50th Eurovision Song Contest was held in Kiev, Ukraine. The winning entry out of 24 finalists was from Greece, with Malta as the runner-up.

The Eurovision Song Contest involves the live television broadcast of popular songs from various European countries. Each country then casts votes for its ten favourites on a 1…12 scale: 12 points for the favorite, 10 for the second favorite, and 8,7,6…1 points in turn for the third to tenth favorite. These votes are based on telephone polls conducted in each country during the broadcast. Votes cast in the 2005 final are shown in Table 1, in the format found on the Eurovision web site (European Broadcasting Union, 2005). Accusations of political influence on the voting patterns have been common, particularly by BBC commentator Terry Wogan (Wikipedia, 2005). A notable example was the failure of any country to assign points to the UK in 2003, possibly in protest against UK involvement in Iraq. Our analysis will confirm that, interpreted using Social Network Analysis techniques, the Contest results do indeed provide a window into European politics.

A difficulty in analysing data from the Eurovision Song Contest has been the enormous variation in the number of participants. The very identity of "Europe" has changed enormously in the past 50 years, and the rules of the Contest have also altered. We avoid these issues by using techniques that allow conclusions to be drawn from a single year's data, thus presenting a "snapshot" of a changing Europe at one point in time. The techniques we use may also be of more general interest.

Figure 1 shows the average vote in 2005 from country *X* to country *Y*, as a function of the total score obtained by country *Y* (an indication of the overall popularity of country *Y*'s entry), and on the distance between countries *X* and *Y* (measured by the number of borders needing to be crossed in order to travel from country *X* to country *Y*, thus eliminating geographical area as a factor). Figure 1 shows that the highest votes generally go to songs whose popularity is shared (i.e. with high total scores), and to songs from nearby countries (i.e. with small distances), presumably because of shared linguistic and cultural factors.

Table 1: Votes Cast for Eurovision Song Contest Final in 2005

			Votes From																																					
		Total Score	Andorra	Albania	Austria	Belarus	Belgium	BosniaHerz	Bulgaria	Croatia	Cyprus	Denmark	Estonia	Finland	France	FYRM	Germany	Greece	Hungary	Iceland	Ireland	Israel	Latvia	Lithuania	Malta	Moldova	Monaco	Netherlands	Norway	Poland	Portugal	Romania	Russia	SerbiaMont	Slovenia	Spain	Sweden	Switzerland	Turkey	Ukraine	UK
Votes To	Greece	230	4	12	4	0	12	6	12	5	12	2	0	3	8	7	12	—	12	2	2	7	0	1	6	4	0	10	4	1	3	10	4	12	2	8	12	7	12	0	12
	Malta	192	0	4	5	5	8	0	0	4	6	10	4	8	7	0	8	8	5	4	10	10	5	2	—	0	5	5	10	0	0	2	12	0	1	7	6	3	8	10	10
	Romania	158	7	5	6	1	7	2	8	0	8	3	0	0	5	2	0	5	10	5	5	12	0	0	7	7	4	3	6	7	12	—	0	3	0	12	2	0	4	0	0
	Israel	154	8	3	1	8	6	0	0	0	0	5	1	5	10	0	5	0	8	0	6	—	2	3	8	6	12	7	5	0	5	7	8	0	0	6	1	1	3	7	7
	Latvia	153	10	0	0	6	5	0	0	7	1	6	10	4	0	0	7	1	1	3	12	6	—	12	10	12	0	0	8	4	6	0	5	0	7	0	3	0	0	1	6
	Moldova	148	0	0	2	7	1	4	6	1	2	0	6	0	2	5	1	7	4	8	0	4	8	10	2	—	0	0	0	3	10	12	10	5	3	4	0	0	7	12	2
	SerbiaMont	137	0	6	12	3	0	10	4	12	10	0	0	0	6	10	3	6	2	0	0	0	1	0	0	1	6	4	0	0	0	6	6	—	10	0	4	12	0	3	0
	Switzerland	128	1	0	0	10	0	0	0	3	4	1	12	10	0	0	4	3	3	7	3	2	12	8	1	0	8	0	3	6	4	0	7	0	6	0	5	—	0	5	0
	Norway	125	2	0	0	4	3	3	1	0	3	12	8	12	0	0	0	4	0	12	4	1	6	5	5	3	0	1	—	8	0	0	0	2	4	3	8	0	0	6	5
	Denmark	125	3	0	0	0	4	0	0	0	0	—	5	2	0	0	6	0	6	10	7	3	4	4	3	0	10	8	12	5	1	4	0	0	0	10	10	0	0	0	8
	Croatia	115	0	2	8	0	0	12	2	—	0	0	2	1	0	8	2	0	7	1	0	0	7	6	0	0	7	2	2	2	0	5	1	10	12	0	0	8	0	8	0
	Hungary	97	6	0	0	2	2	1	5	6	7	0	3	0	3	1	0	2	—	6	0	8	3	0	0	0	0	0	0	10	2	8	3	6	0	5	0	0	6	2	0
	Turkey	92	0	8	7	0	10	8	3	0	0	8	0	0	12	4	10	0	0	0	0	0	0	0	0	0	0	12	0	0	0	3	0	0	0	0	0	6	—	0	1
	BosniaHerz	79	0	0	10	0	0	—	0	10	0	4	0	0	0	3	0	0	0	0	1	0	0	0	0	0	0	6	7	0	0	0	0	4	8	0	7	5	10	0	4
	Russia	57	0	0	0	12	0	0	0	0	0	0	7	7	0	0	0	0	0	0	0	0	10	7	0	10	0	0	0	0	0	0	—	0	0	0	0	0	0	4	0
	Albania	53	0	—	3	0	0	5	0	2	0	0	0	0	1	12	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	10	2	0	0
	FYRM	52	0	10	0	0	0	7	7	8	0	0	0	0	0	—	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	7	5	0	0	2	5	0	0
	Cyprus	46	0	7	0	0	0	0	10	0	—	0	0	0	0	0	0	12	0	0	0	0	0	0	12	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	3
	Sweden	30	0	0	0	0	0	0	0	0	0	7	0	6	0	6	0	0	0	0	0	0	0	0	0	5	3	0	1	0	0	0	0	0	0	2	—	0	0	0	0
	Ukraine	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	12	7	0	2	0	0	1	0	0	0	—	0
	Spain	28	12	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	—	0	4	0	0	0
	UK	18	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	8	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	—
	France	11	5	1	0	0	0	0	0	0	0	0	0	0	—	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	Germany	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	—	0	0	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0

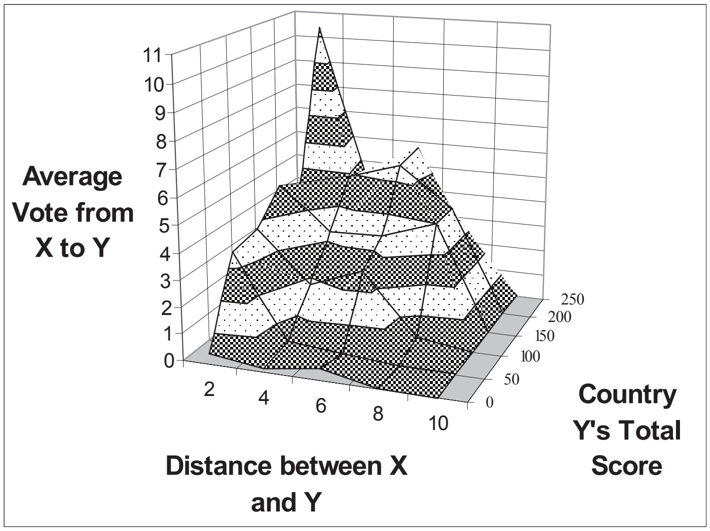


Figure 1: Eurovision Song Contest Votes from Country *X* to Country *Y* as a Function of Total Score and Distance Between Countries

Although the Eurovision Song Contest is ostensibly a competition on song quality, we can adjust for song quality to obtain a "friendship" network, similar in structure to friendship networks between people. The resulting social network has *valued links*: a high vote for an otherwise unpopular song indicates maximum friendship, while a low vote for a popular song indicates least friendship. We analyse the friendship network using techniques previously developed for valued networks (Dekker, 2005), which combine network-analysis methods with statistical methods. Statistical techniques for Social Network Analysis are also discussed by Wasserman and Faust (1994), but the methods they present have limited utility for valued networks.

Our analysis of the Eurovision Song Contest data reveals a set of *friendship blocs*, and a significant tendency to vote for nearby countries. Some individual countries have more unusual voting behaviour, and we briefly discuss reasons for this. Finally, we compare our results with past studies of the Eurovision Song Contest.

RESULTS

The Friendship Network

We can measure the quality (or at least popularity) of country *Y*'s song performance by using country *Y*'s total score S_Y. This total provides a measure of how highly that country's song was rated by a Europe-wide audience. When we plot the vote V_{XY} from country *X* to country *Y* against country *Y*'s total score S_Y, we obtain a weak linear relationship, shown in Figure 2. The line of best fit was:

$$V_{XY} \approx 0.026\, S_Y \qquad \text{(Figure 1)}$$

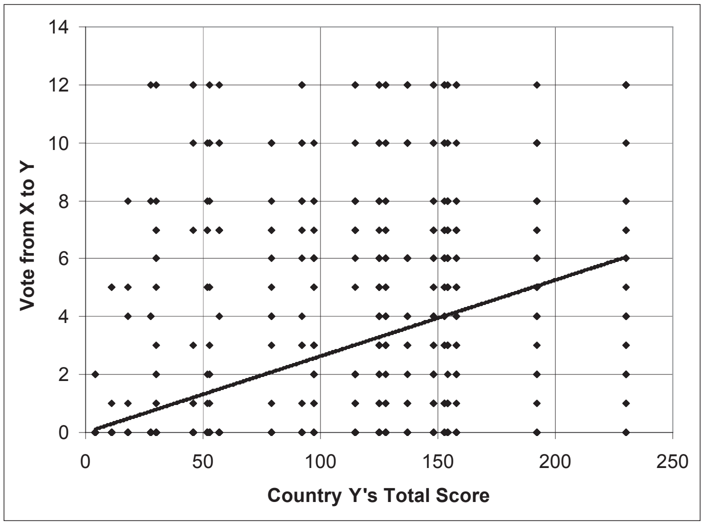


Figure 2: Eurovision Song Contest Votes as a Function of Total Score

The correlation here is a weak 0.44 (r^2 = 20%), but is statistically extremely significant ($p < 10^{-44}$), i.e. votes are indeed partially determined by the shared perception of song quality, as we would expect. We can therefore adjust scores for song characteristics by subtracting the predicted vote from the actual vote, giving a friendship score F_{XY} (we also add 6.1 to ensure that the result is positive, in the range 0 to 17.4):

$F_{XY} = 6.1 + V_{XY} - 0.026\, S_Y$ (Figure 2)

Having subtracted the shared perception of song "quality" from the votes, the numbers F_{XY} which remain provide an indication of the bias that country *X* has towards country *Y*. These numbers form a social network with a structure similar to that obtained by asking a group of people how much they like each other, and we therefore refer to it as a friendship network, in the sense that countries like Norway and Denmark can be informally described as "friends." However, the biases between countries are naturally more complex than friendship between individuals, being influenced by cultural, political, and other factors.

Bruine de Bruin (2005) and Haan, et al. (2005) have found that, as well as song quality, the order of performance also determines country *Y*'s total score S_Y. However, by adjusting for S_Y, we are also compensating for that factor.

Figure 3 shows the friendship network, laid out using Spring Embedding (Freeman, 2000), a process equivalent to Multidimensional Scaling (Brandes, 2001). For clarity, Figure 3 shows only arrows corresponding to votes with a high friendship score $F_{XY} > 12$. In this network, friendship tends not to be returned: the correlation between F_{XY} and its inverse F_{YX} is a weak 0.40 (we will discuss the reasons for this later in the paper). In our experience, a correlation of 0.6 or more would indicate a symmetric relationship (Dekker, 2005). A consequence of the lack of symmetry in friendship is that the concept of *link distance* between people which we introduced in previous work (Dekker, 2005) must be used with caution, since it is based on symmetrical relationships. Figure 3 should therefore be interpreted with care, particularly since the correlation between friendship scores F_{XY} and distances in the diagram is a relatively weak – 0.38.

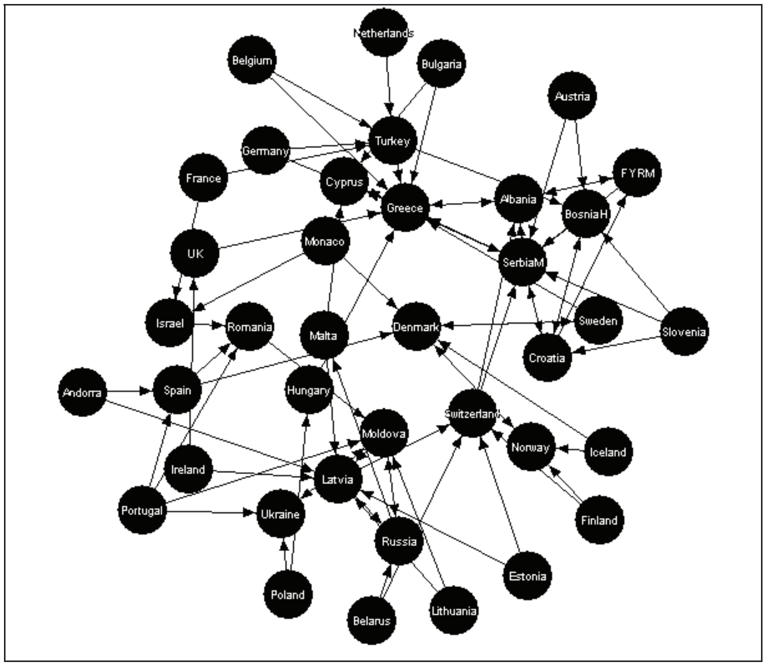


Figure 3: Eurovision Song Contest Friendship Network

Conclusions about individual countries from distances in Figure 3 should be made with care. However, Figure 3 does contain several visually apparent *friendship blocs*, composed of nearby countries which vote for each other. Figure 4 shows more clearly these friendship blocs, which are:

clearly these friendship blocs, which are:

- **Eastern**: former USSR, Romania, Hungary, Poland.
- **Nordic**: Norway, Sweden, Denmark, Finland, Iceland.
- **Balkan**: former Yugoslavia, Albania.
- **Eastern Mediterranean**: Greece, Cyprus, Malta, Bulgaria, Turkey.
- **Western**: other countries.

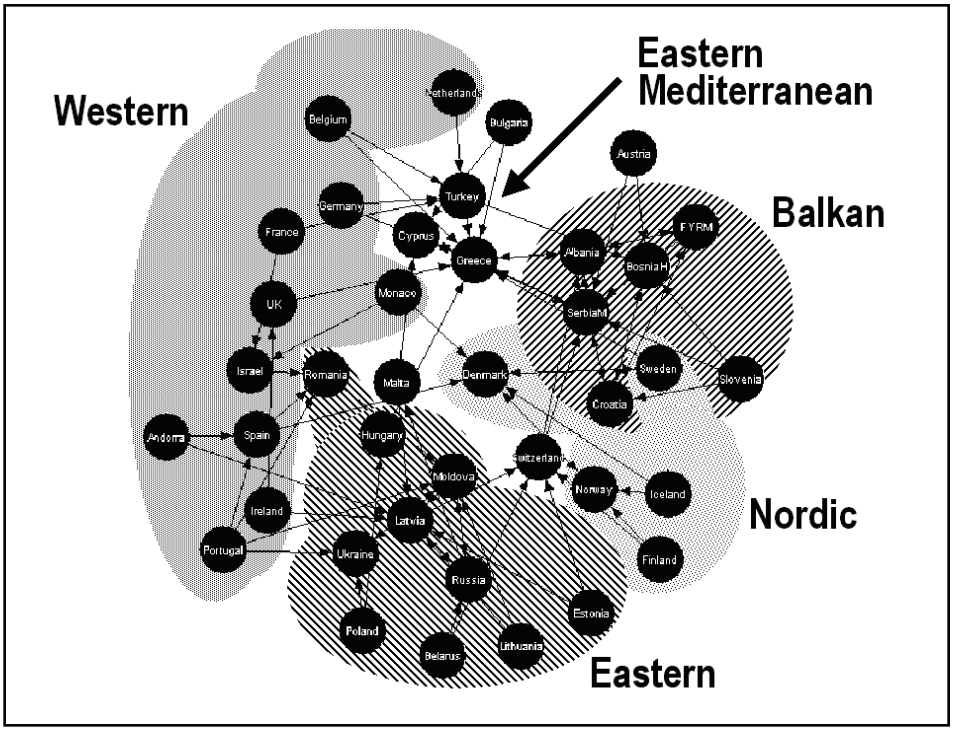


Figure 4: Friendship Blocs in the Eurovision Song Contest

We also grouped countries computationally, by considering only votes with a high friendship score $F_{XY} > 12$, and applying the Strongly Connected Components algorithm (Gibbons, 1985). This algorithm has the advantage of being fully deterministic, and finds three clusters in this case. They are subsets of the Eastern bloc (Russia, Latvia, Moldova, Ukraine), Nordic bloc (Norway, Sweden, Denmark), and Balkan and Eastern Mediterranean blocs together (Greece, Cyprus, Albania, and former Yugoslavia). The Strongly Connected Components algorithm groups Malta with the Eastern bloc because its second-highest vote was for Latvia, while Malta received Russia's highest vote. However, it seems more appropriate to define the Eastern bloc as the former Warsaw Pact countries (with the exception of Bulgaria, which gave its highest votes to Greece and Cyprus, and is therefore grouped with them). It also seems appropriate to separate the Balkan and Eastern Mediterranean blocs, which are visibly distinct in Figure 3.

Grouping countries using a Simulated Annealing algorithm (Hecht-Nielsen, 1990) gives different results each time the algorithm is run, but consistently separates the Balkan and Eastern Mediterranean blocs, while giving inconsistent groupings for the other countries. This supports our separation of the Balkan and Eastern Mediterranean blocs.

Another common way of grouping countries is the use of taxonomic trees, as in Fenn at al. (2005). However, taxonomic trees are known to be sensitive to random noise in the data. We applied a taxonomic algorithm, which randomly alters all friendship scores by between 0 and 0.1%, calculates taxonomic trees using neighbour-joining (Pachter and Sturmfels, 2005), repeats this 100 times, and then takes only the relationships common to all 100 trees. There were very few of such common relationships, underscoring the brittleness of taxonomic-tree formation. Tree relationships which were not common, i.e. which altered in the face of only a 0.1% alteration in the data, are clearly due to chance, and therefore meaningless. Figure 5 shows the common relationships. The only cluster recognisable in this diagram is the Eastern bloc, not including Hungary and Romania (which are closer to the Western bloc) or Estonia (which is closer to the Nordic bloc).

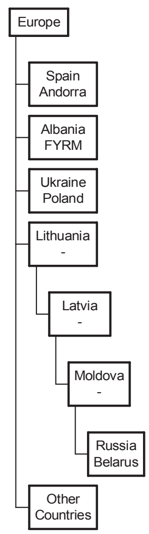


Figure 5: Common Relationships in Taxonomic Trees for Countries

As would be expected, friendship scores within a bloc were higher than between blocs (on average, 8.1 versus 5.6, significant at $p < 10^{-20}$). Table 2 shows the average friendship scores between and within blocs. Interestingly, the Western bloc was the least cohesive: within-bloc scores were lowest for the Western bloc.

Table 2: Average Friendship Scores Within and Between Blocs

		To				
		Eastern	Nordic	Balkan	East Med	Western
From	Eastern	**7.9**	6.1	5.0	4.4	6.5
	Nordic	5.3	**11.3**	4.8	5.9	6.0
	Balkan	4.9	4.5	**11.0**	5.8	5.0
	East Med	6.2	4.7	6.0	**8.7**	5.6
	Western	5.7	5.8	5.5	6.8	**6.6**

Figure 6 shows the above-average between-bloc scores. The pairs Balkan/Eastern Mediterranean, Western/Eastern, Western/Nordic, and Western/Eastern Mediterranean had above-average scores in both directions, and there was also a triangle of unidirectional above-average scores. Figure 6 highlights the position of the Balkan countries as “new arrivals” in Europe, and a possible “bridging” role played by the Eastern Mediterranean countries. Individual Western countries such as Austria and Switzerland may also play a “bridging” role. The links between the Western bloc and the other blocs reflects the past dominance of Europe (and the Eurovision Song Contest) by the Western bloc.

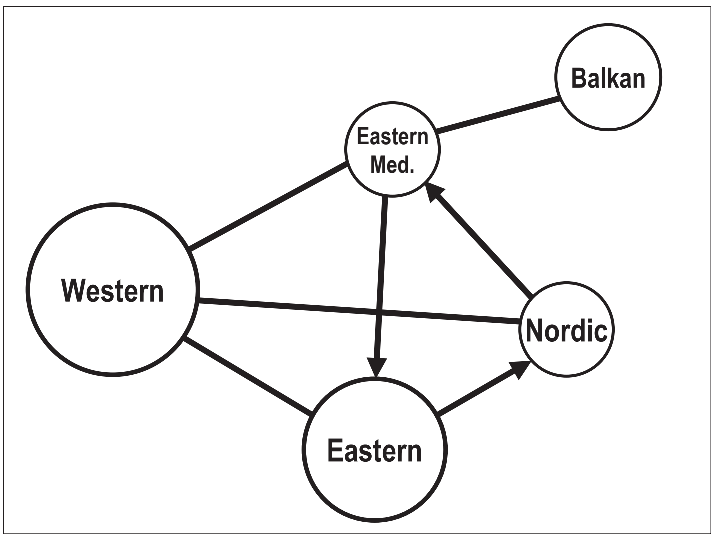


Figure 6: Links Between Friendship Blocs in the Eurovision Song Contest

Austria and Switzerland are somewhat exceptional countries. Both gave strong votes to Balkan countries, presumably because of large numbers of Balkan immigrants (Switzerland also received strong votes from Finland and the Eastern bloc). Similarly, Romanian immigrants in Spain seem to have given strong votes to their home country, as did Turkish immigrants in the five countries with the most Turkish immigrants: Belgium, France, Germany, Austria, and the Netherlands (Manço, 2004). This kind of voting was in general not returned. Table 3 lists the fifteen greatest asymmetries in friendship scores, where the score F_{XY} was 7 or more points greater than the reverse score F_{YX}. Of these, six can be tentatively attributed to immigrants (Niessen et al., 2005). However, this attribution cannot, of course, be confirmed without surveying Eurovision voters on the reasons for their vote.

Table 3: Fifteen Greatest Asymmetries in Friendship Scores Between Countries

Country Pair		Friendship Score	Reverse Score	Possible Reason
Latvia	Switzerland	14.7	2.1	?
Switzerland	Albania	14.7	2.7	?
Switzerland	Serbia/Montenegro	14.5	2.7	Immigrants
Malta	Cyprus	16.9	7	?
France	Turkey	15.7	5.8	Immigrants
Denmark	Turkey	11.7	2.8	Immigrants
Spain	Romania	13.9	5.4	Immigrants
Russia	Malta	13	4.6	?
Turkey	Greece	12	3.7	?
Moldova	Sweden	10.3	2.2	?
Germany	Turkey	13.7	6	Immigrants
Latvia	Russia	14.6	7.1	Former USSR
Spain	Denmark	12.8	5.4	?
Albania	Cyprus	11.9	4.7	?
Switzerland	Turkey	9.7	2.7	Immigrants

The Effect of Distance

Although we are not able to use the concept of link distance (Dekker, 2005) for analysis, we can use related statistical techniques to examine the factors that determine friendship scores. A good predictor of the friendship score F_{XY} was the distance D_{XY} between countries which we discussed above (measured by the number of borders needing to be crossed in order to travel from country X to country Y). This is presumably because of cultural and linguistic factors shared between nearby countries. Economic factors (as measured by differences between country's GDPs) did not seem to have an effect, nor did population size. Linguistic difference alone (as measured by distance in a four-level language family tree) had a small effect ($r^2 = 2\%$), but this effect vanished when D_{XY} was included, since D_{XY} already incorporates cultural and linguistic factors. The line of best fit was:

$$F_{XY} \approx 7.8 - 0.46\, D_{XY} \qquad (3)$$

The correlation here was a very weak –0.24 ($r^2 = 6\%$), but was statistically extremely significant ($p < 10^{-12}$), thus providing additional justification for defining blocs of countries which are geographically close to each other. Figure 7 shows the relationship graphically. An alternative way of describing the result is that the original vote V_{XY} from country X to country Y (see Figure 1) can be explained (by analysis of variance) as 20% due to country Y's total score S_Y, 4% due to the distance D_{XY} between countries, and 76% due to other factors. We would expect some of the 76% to be explained by the numbers of immigrants from country Y living in country X, but accurate statistics on this are difficult to obtain.

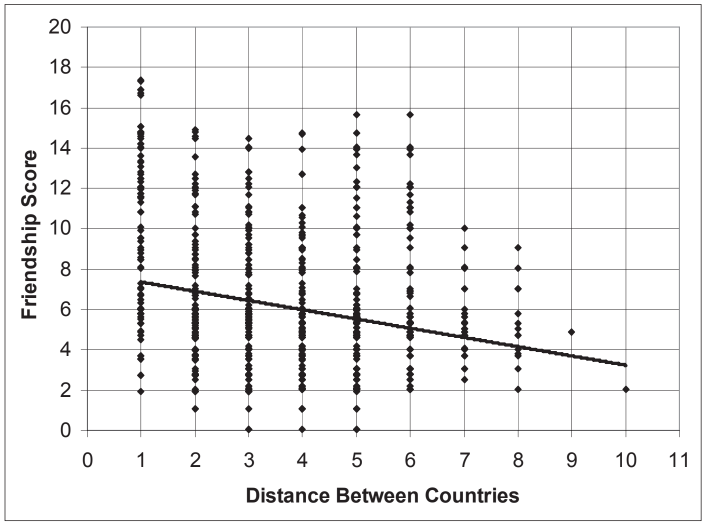


Figure 7: Friendship Scores as a Function of Distance between Countries

Centrality

Various forms of *centrality* concept have shown great utility in Social Network Analysis (Wasserman and Faust, 1994). We therefore calculated *valued centrality* (Dekker, 2005) scores for each of the countries in Figure 3. This centrality measure takes "closeness" to be the inverse of distances, d_{ij}, along network paths, and obtains valued centrality, C_i, by averaging closeness values:

$$C_i = \frac{1}{n-1} \sum_{j \neq i} d_{ij} \quad (4)$$

This is the most suitable definition of centrality for valued networks (Dekker, 2005). However, since measuring distance along network paths is not really appropriate with non-symmetric friendship relationships, the valued centrality scores should be interpreted with some caution. There was a strong correlation of 0.84 between valued centrality scores and total Contest scores (r^2 = 70%, significant at $p < 10^{-6}$), as shown in Figure 8.

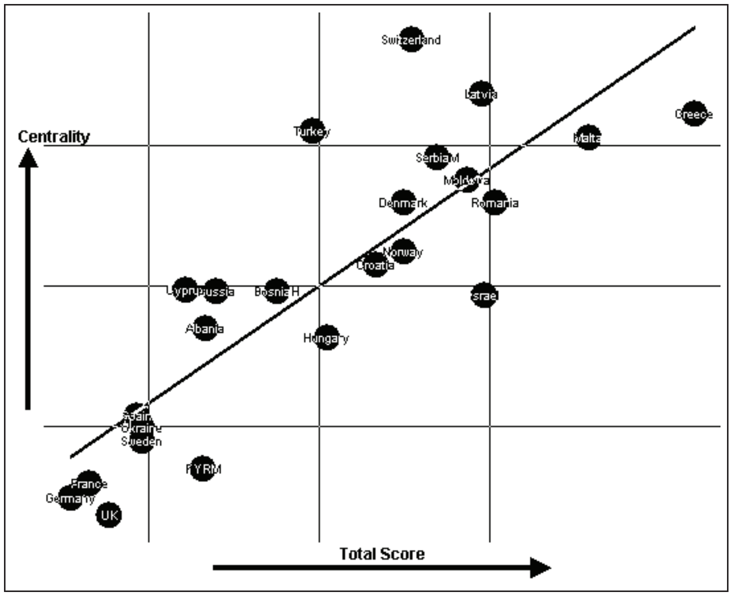


Figure 8: Valued Centrality Scores as a Function of Total Contest Score

Of particular interest are the countries whose valued centrality is higher than their total score would indicate, in particular Turkey and Switzerland. This seems to reflect the high votes which Turkey obtained from Turkish immigrants in Western countries, and "neutral" Switzerland acting as a "bridge" between blocs by giving votes to the Balkans while receiving them from the Eastern bloc.

DISCUSSION

By adjusting Eurovision Song Contest votes to compensate for song quality, we have obtained a friendship network, which can indeed "reveal by homology the structure of political Europe" (le Guern, 2002). The structure that our analysis has revealed is very different from that reported by Yair (1995), who partially compensated for song quality by averaging votes over several years. Yair's study revealed the Western bloc as dominant. In contrast, today's Europe is very different, with Western countries being least central and least cohesive, and Central Europe being more important.

In general, we found that friendship scores were highest for nearby countries, resulting in five friendship blocs: Western, Eastern, Nordic, Balkan, and Eastern Mediterranean. The new countries of the Balkans were most isolated, with the Eastern Mediterranean countries, Austria, and Switzerland acting as "bridges." The past dominance of the Western bloc is reflected in its close ties to the other blocs (excluding the Balkans).

Recent work by Fenn, et al. (2005) examines Eurovision Song Contest data from the period 1992–2003, using the framework of complex dynamical networks, and also averaging votes over several years. For the smaller set of countries competing in that period, they also found regional clustering, particularly the pair Greece/Cyprus, the Nordic bloc (including Estonia), and the Western bloc, which (unlike Yair) they did not find to be cohesive. However, this analysis was based on taxonomic trees, which may be deceptive. Fenn et al. also found UK voting to be consistently the most in tune with the rest of Europe—and indeed, in 2005, the UK was the only country which gave its top two votes to the ultimate winner (Greece) and runner-up (Malta). This may be a result of the UK's reduced involvement in regional ties and/or conflicts.

Gatherer (2006) examines data in five-year windows over several years (1975–2005), comparing them against simulation results, which allows him to perform tests of statistical significance on links. He finds two large clusters in the period from 2001 to 2005: a "Balkan Bloc," which includes our Eastern Mediterranean cluster, and a "Viking Empire," of the Nordic countries with Latvia, Lithuania, and Estonia. However, his analysis does not allow relationships outside these clusters, or relationships between clusters, to be inferred. In addition, this and other previous studies do not adequately eliminate song quality as a factor, and therefore potentially obscure the true bias or friendship relationships between countries.

The high valued centrality score for Turkey in our study emphasises the importance of immigrants from one country living in another. Van der Veen (2002) points out that such expatriate workers are also of importance in the formation of a pan-European social identity. The high valued centrality score for Switzerland suggests that it plays a "bridging" role within the new Europe, although the reasons for this are not completely clear.

Just as the mathematical techniques we have presented have illuminated the structure of Europe, they can also be used to generate friendship networks from other competitions where the participants (or their representatives) double as the voting jury. For example, in the judging of the Olympic Games, the difference between individual judge's scores and the average for a particular performance can be used as a friendship vote from the judge's country to the athlete's. Unlike previous techniques for studying the Eurovision Song Contest, the method we describe does not require many years of historical data, and hence can be used for studying social structures which are in a state of flux. The analysis techniques presented in this paper can also be used for analysing other friendship networks, as well as trust networks, which have a similar structure.

References

Brandes, U. (2001). Drawing on Physical Analogies. In *Drawing Graphs: Methods and Models* (Michael Kaufmann and Dorothea Wagner, eds), *Springer-Verlag Lecture Notes in Computer Science* **2025:** 71–86.

Bruine de Bruin, W. (2005). Save the last dance for me: unwanted serial position effects in jury evaluations. *Acta Psychologica* **118**: 245–260.

Dekker, A.H. (2005). Conceptual Distance in Social Network Analysis. *Journal of Social Structure*, **6** (3): www.cmu.edu/joss/content/articles/volume6/dekker/

European Broadcasting Union, (2005). Eurovision Song Contest web site, accessed 10 June 2005: www.eurovision.tv

Fenn, D., Suleman, O., Efstathiou, J., and Johnson, N.F. (2005). How does Europe Make Its Mind Up? Connections, cliques, and compatibility between countries in the Eurovision Song Contest. Oxford: Oxford University: arxiv.org/abs/physics/0505071

Freeman, L.C. (2000). Visualizing Social Networks. *Journal of Social Structure*, **1**(1): www.cmu.edu/joss/content/articles/volume1/Freeman.html

Gatherer, D. (2006). Comparison of Eurovision Song Contest Simulation with Actual Results Reveals Shifting Patterns of Collusive Voting Alliances. *Journal of Artificial Societies and Social Simulation*, **9**(2): jasss.soc.surrey.ac.uk/9/2/1.html

Gibbons, A. (1985). *Algorithmic Graph Theory*. Cambridge: Cambridge University Press.

Haan, M., Dijkstra, S., and Dijkstra, P. (2005). Expert Judgment Versus Public Opinion – Evidence from the Eurovision Song Contest. *Journal of Cultural Economics* **29**: 59–78.

Hecht-Nielsen, R. (1990). *Neurocomputing*. Reading, Massachusetts: Addison Wesley.

Niessen, J., Schibel, Y., and Thompson, C. (eds.), (2005). Current Immigration Debates in Europe. Brussels: Migration Policy Group: www.migpolgroup.com/documents/3055.html

le Guern, P. (2002). From National Pride to Global Kitsch: the Eurovision Song Contest. Lille: University of Lille: wjfms.ncl.ac.uk/leguWJ.htm

Mançо, U. (2004). Turks in Western Europe. Brussels: Centre d'Etudes Sociologiques, Facultes Universitaires Saint-Louis: www.flwi.ugent.be/cie/umanco/umanco3.htm

Pachter, L. and Sturmfels, B. (2005). *Algebraic Statistics for Computational Biology*. Cambridge: Cambridge University Press.

van der Veen, A.M. (2002). Determinants and Implications of European Identity: An Investigation Using Eurobarometer Data. Philadelphia, PA: University of Pennsylvania: www.ssc.upenn.edu/~maurits/papers/EUsupport.doc

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Wikipedia, (2005). Eurovision Song Contest, accessed 10 June 2005: en.wikipedia.org/wiki/Eurovision_Song_Contest

Yair, G. (1995). 'Unite Unite Europe' The political and cultural structures of Europe as reflected in the Eurovision Song Contest. *Social Networks*, **17**(2): 147–161.

Where Does Help Come From: A Case Study of Network Analysis in an Academic Group

Pengxiang Li, Youmin Xi, and Xiaotao Yao
School of Management, Xi'an Jiaotong University, Xi'an, China [1]

This paper explores the relationship between the transferring activities of such resources as information, knowledge, social support, and tie strength of interaction among individuals in an academic group. We investigated the academic group led by Professor Xi in the School of Management of Xi'an Jiaotong University through questionnaire. The empirical data for four kinds of networks such as help, communication, friend, and research cooperation were collected and analyzed. Findings show that the interaction ties in the group cannot be viewed simply as strong ties, but the complex situation with strong and weak ties interweaved. Among the activities of transferring resources, nearly sixty percent of help came from strong ties and about sixty-five percent of help providers were the core actors in the group. This means that strong ties are the main channels of resource transfer while the role of weak ties should still be given attention. The strong ties with the core actors in the group can make it easier to get more valuable help than weak ties. However, new information and possible opportunities are more likely to be provided through weak ties than strong ties.

This work is supported partially by the Excellent Innovative Research Group Funds grant No. 70121001 from the National Science Foundation in China and partially by the National Science Foundation grant No.70571062 in China.

We would like to thank all of the group members for finishing our questionnaire to support our network analysis research in spite of busy work. We are very proud of them. Meanwhile we wish to express our gratitude to all the teachers and graduates who gave us much helpful suggestions at the seminar of the Excellent Innovative Research Group sponsored by Management school, Xi'an Jiaotong University. Finally, thanks are due to two anonymous reviewers for their helpful suggestions.

INTRODUCTION

Academic groups are led by supervisors in universities or institutes of scientific research. Among these groups, members engage in research work with their supervisors, writing papers and their dissertations. Every year some newcomers join such groups and some old members leave after graduation. These groups develop and refresh year by year, and the research work continues. In the process of their research and studies, the transferring activities of resources such as information, knowledge and social support take place very often. The resource providers exchange their information, knowledge, or social support for others with the seekers in such ways as telephone, e-mail, network platform or private conversation.

Specifically, there are three definitions of "help" referred to in this paper. The first one is providing such information as academic meetings, literature searches, software downloads, paper submitting, and business recruitment for other actors. This kind of help is available for all actors who have a weak tie with providers because of less cost. The second one is the exchanges of some knowledge about how to do experiments or research designs, and some existing results such as experiment data, questionnaire tables, program modules, and so on. These resources are only transferred between two actors who have much trust for each other. The third level of meaning refers to the social support such as acquiring understanding, comfort, and encouragement from other people when a person comes across difficulties or something unpleasant. The transfer of these resources require especially close relationships between seekers and providers.

Except for the shared relationship between graduates and their supervisor, the interactions among group members include the following four aspects: helping each other in their study and research work, communicating with each other, private friendship, and research collaboration. These interactions can be classified into two types of ties in terms of tie strength: (a) acquaintanceship (weak tie), knowing each other as a member of the group and (b) friendship (strong tie), friendship gradually formed during routine academic activities, cooperation, and long-term private connections. Our problems here are whether the help benefiting from information, knowledge and social support comes from strong ties or weak ties, and whether the providers are core actors or periphery actors in their social networks.

In the related social network literatures, there have been many studies on the role in which strong ties and weak ties play in transferring resources. The management problems involved in these studies are sharing knowledge across organization subunits (Hansen, 1999), the transfer of knowledge (Levin et al., 2002), organizational conflict (Nelson, 1989), communication (Pickering and King, 1995), technical advice (Constant et al., 1996), employment or finding a job (Montgomery, 1994, Bian, 1997, Bian and Ang, 1997), interfirm exchange relations (Keister, 1999), buyer's selection (Kiecher and Hartman, 1994), and so on. The differentiation of strong and weak ties in these studies is according to the definition made by Granovetter (1973, 1982), which is based on four dimensions: time, emotional intensity, mutual confidence, and reciprocity. Strong ties are maintained through frequent and emotionally intense communication, often entailing the sharing of confidences, and over time, the establishment of reciprocity between the parties. Weak ties are maintained through less frequent and less emotionally intense communication, in relationships that do not require or encourage sharing of confidences or establishment of strong reciprocities.

In these literatures, the intragroup (or intraunit) interaction was usually viewed as strong ties, while the intergroup (or interunit) interaction as weak ties. In fact, the role of weak ties was only focused on in the exploration of computer-mediated communication (Pickering and King, 1995) and the usefulness of electronic weak ties for technical advice (Constant et al., 1996). Krackhardt (1992) defined the relationship of "philos" as the criterion of strong ties. In the case of Silicon System, almost all actors have strong ties with others. His paper emphasized the role of strong ties in the organization reform.

However, there has been little attention on the role that strong ties and weak ties play in resource transfer inside a group. Flache (2002) studied group solidarity in a highly cohesive group of rational agents to emphasize the weakness of strong ties, but his work was not a case study. It seems that intragroup interactions can be always approximately viewed as strong ties and the resources transferred among group members are mainly some information or knowledge that everyone knows. In the literatures on academic groups, there has been no network case study that has discussed the role of strong ties and weak ties in resource transfer. The research sites in existing literatures are manager groups (Krackhardt, 1987; 1992; Krackhardt and Jeffrey 1993; Barsky, 1999), work groups, departments or whole organizations (Brass and Burkhardt, 1992).

"The strength of weak ties" (Granovetter, 1973) and "the strength of strong ties" (Bian, 1997) sound a little contradictive but with much reason. Both of the above hypotheses have much empirical evidence to support. This arouses our research interest of knowing what interactions exist in an academic group and how such resources as information, knowledge, and social support transfer in it. Perhaps strong ties play a more important role than weak ties in the process of resource transfer, or just the opposite. We wonder which part of the members in their networks help the others, and which position are the providers situated in the networks. The motivation for this case study is to illustrate the following points: (1) all the intragroup interaction ties are not always strong ties, but both strong ties and weak ties; (2) resources transferred among group members consist of individuals knowledge but also what each actor continually learns from outside the network, giving, groups many new things to exchange. The case study presented in this paper has three distinctions. Firstly, existing research focuses on resource transfer between one group and another, or between the inside and the outside. Our work concentrates on the role which strong ties and weak ties play in one group. Secondly, our research site is the academic group led by its supervisor. This kind of group is the informal group oriented at the accumulation and creation of knowledge, in which some graduated students leave and newcomers enroll every year. This is different from workgroups and manager circles. Thirdly, the case study not only examined the role in which tie strength played in getting help, but also analyzed the position in which help providers lie in social networks of the group.

The reminder of this paper is organized as follows. Section 2 introduces how the data collection was done and how to process the data to acquire the social networks of the group. The purpose of section 3 is to illustrate that group members are more likely to get help through strong ties than weak ties. Core and peripheral structure analysis is conducted by using network data in section 4, which aims at examining which position help providers lie in the network structure. Our work will demonstrate that core members in social networks of the group are likely to provide more help than peripheral members. Section 5 explains why the intragroup interaction ties have both strong ties and weak ties and why some people are core members, semi-peripheral, and peripheral ones. The final section is the conclusion of the paper.

METHODS

The help network transferring resources among group members is a directed network, which edges are pointed at the providers from the receivers in the network. This data is collected by each member identifying the person who provided help for him (her). The first question in the questionnaire asked for the person(s) who provided help in the activities related to research work. Hence, the network data is the result of how group members subjectively perceive each other.

Friendship is denoted as strong ties among group members in this paper. Whether or not a tie exists between one group member and another is decided according to his or her own judgment. The strength of this kind of tie exceeds the ordinary acquaintanceship in all three dimensions such as frequency of getting in touch with, duration of affiliating with, and perceived trustworthiness. The next question in questionnaire asks which persons thought of as friends in the group. Different people may have different meanings for friendship. While some have strict standards of judgment, others have less but they all belong to the range of strong ties compared with acquaintanceship; however, these differences have no influence on our final results.

Communication network as a measurement of tie strength reflects the frequency of getting in touch with each other during research work. We simply ask how often the group members get in touch with each other. The three answer choices are as follows: (1) meet often and communicate by telephone or e-mail, (2) communicate only by telephone or e-mail, and (3) almost disconnected. The first one represents a strong tie with the most frequent communication in the above cases. The third one is weak tie. Only the second choice has a little bit of complexity. Maybe they are friends with the lower frequency of communication only during the investigation.

Perhaps they just have acquaintanceship, which is related to the history and background of their affiliations before.

The purpose of investigating the cooperation network during research work is to give us a further validation of strong ties. At the same time, the investigation also reflects which person is the most important one in the group. Obviously, people who are willing to cooperate with others may have certain specialties in research work. The third question asks with whom you cooperate (presently or in the future) when writing a textbook, publishing papers or other works.

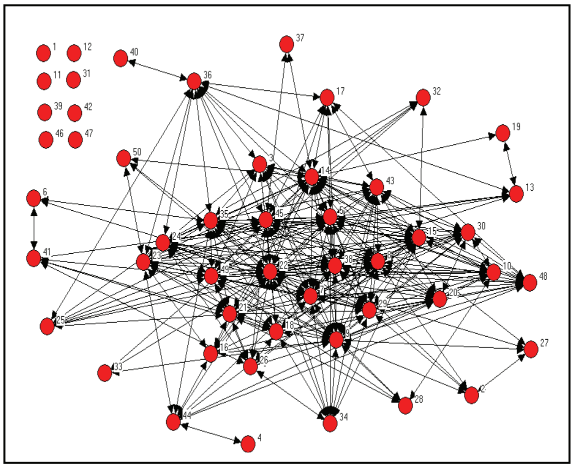
In addition, the questionnaire includes some attribute information of the group members such as name, research interest, scholar degree and post, enrollment time, and so on. Each question attaches a roster of the whole group to choose by responders. Appendix 1 is the attribute variables table of members in the academic group.

Based on the address list of the group members, we sent 50 copies of the questionnaire and received 37 copies. We code all of the group members to substitute for their true name by using random number table (McClave and Terry, 2000, p797-799). Each member's response was coded as a binary variable (i.e. presence of relationship) for questions 1 to 3, coded as valued network (i.e. number of 2,1,0 corresponding to the three choices respectively) for question 4, and then entered into a number-by-number adjacency matrix.

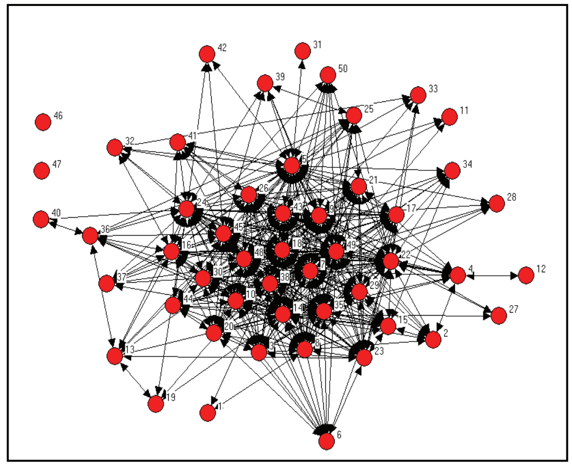
Among the 13 group members who had not returned their questionnaire, there were 12 persons we called "nominal members". They seldom took part in the academic activities organized by the group. The interactions between these nominal members and the other people in the group were found through the questionnaires returned. The last unreturned questionnaire was from the supervisor, who is the leader of the group. The information provided by returned questionnaires was enough to show the closeness of ties between the supervisor and the group members.

RESULTS

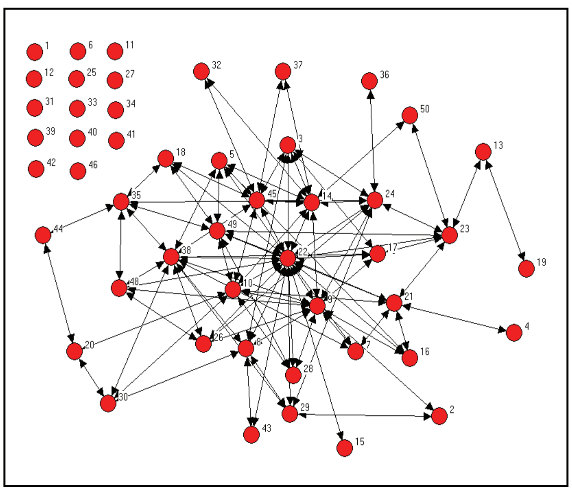
The four matrixes were symmetrized via the maximum method (i.e. if x_{ij}=1 or x_{ji}=1, then x_{ij}= x_{ji}=1), and this can be performed by Ucinet 6 software (Borgatti et al., 2002). Two kinds of matrixes (symmetrized and nonsymmetrized) were saved at the same time because of the following reasons: (1) sometimes the research questions address group issues (rather than relationships between specific pairs); (2) The relationship such as cooperation and communication are symmetric themselves, and friendship seems to be symmetric to some extent. (3) There are 13 lines missing data in the four matrixes because 13 copies of questionnaire were not returned, but the corresponding 13 columns in the matrixes contain valuable information. Symmetrization can make up this shortage. We need to use nonsymmetrized networks when we pay attention to the ties between specific pairs.



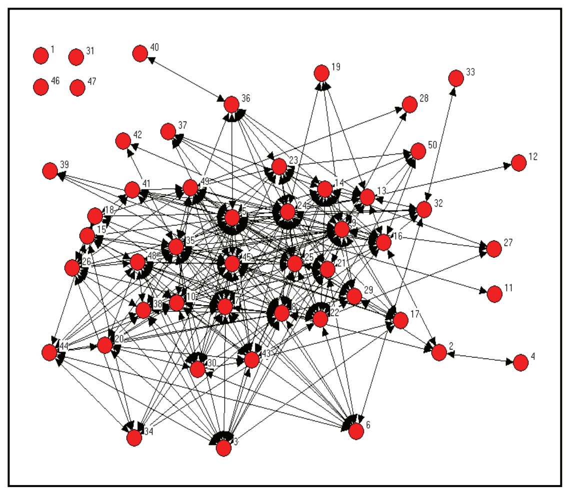
(a) Help Network



(b) Communication Network



(c) Collaboration Network



(d) Friend Network

Figure 1 Four Kinds of Networks of the Academic Group after symmetrization. The isolated nodes are put on the top left corner. The labels of nodes are corresponding to the attribute information table attached in appendix 1.

Help comes mainly from strong ties

We denote the adjacency matrixes of the four kinds of networks respectively as H= $\{h_{ij}\}$ for the help network, F= $\{f_{ij}\}$ for the friend network, R= $\{r_{ij}\}$ for the research collaboration network, and C= $\{c_{ij}\}$ for the communication network. Both the help and the friend networks are directed networks, so the data that we used for these two networks is the original data, which was not symmetrized. While the communication and collaboration networks are undirected networks, we use the symmetrized data. According to the definition of interaction ties in the networks, the meanings of elements in the adjacency matrixes are listed in table 1.

Table 1 Meanings of elements in the adjacency matrixes

Network	Possible Value	Meanings of Elements in the Adjacency Matrixes
Friend Network $F=\{f_{ij}\}$, $(f_{ij}\neq f_{ji})$	$f_{ij}=0$	Actor j is thought not to be one of his friends by actor i.
	$f_{ij}=1$	Actor j is thought to be one of his friends by actor i.
Help Nework $H=\{h_{ij}\}$, $(h_{ij}\neq h_{ji})$	$h_{ij}=0$	Actor i thought actor j provided no help for him.
	$h_{ij}=1$	Actor i thought actor j did provide some help for him.
Communication Network $C=\{c_{ij}\}$, $(c_{ij}=c_{ji})$	$c_{ij}=0$	Two actors almost disconnected.
	$c_{ij}=1$	Two actors met little and communicated only by telephone and e-mail
	$c_{ij}=2$	Two actors met often and communicated by telephone and e-mail
Collaboration Network $R=\{r_{ij}\}$,$(r_{ij}=r_{ji})$	$r_{ij}=0$	It is impossible for two actors to coauthor.
	$r_{ij}=1$	There is a coauthor tie between two actors.

Comparison results among the above four kinds of networks show that using friendship as the approximate criterion of strong ties is reasonable. If two actors are friends or at least one considers unilaterally the other as his friend, they will be more likely to collaborate, help each other, and communicate more frequently than if they are not. For example, the proportion of communicating at least by telephone and e-mail between friends is nearly 96%, 24% of that proportion is between two persons with weak ties; the proportion of coauthorship between friends is 28.5%, while 2.5% of that proportion is between two persons with weak ties. The other results of network comparison are shown in figure 2.

It is shown from the results of network comparison that nearly sixty percent of help comes from strong ties among the activities of transferring resources. Strong ties are the main channels of resource transfer but with only a slight advantage. The role of weak ties should still be given attention because forty percent of help comes from weak ties. The interaction ties in the group cannot be viewed simply as strong ties but the complex situation with strong and weak ties interweaved. Strong ties can make it easier to get useful knowledge and affectional comfort that is impossible to gain through weak ties. However, new information and possible job opportunities may be provided through the weak ties. Such information and opportunity cannot be gained through strong ties.

Most of help providers are core actors

The network structure of the academic group is actually a big cluster in which four kinds of interaction ties are interweaved together. The central part consists of core actors that have cohesive linkages with the others. The outside of the cluster is composed of some ordinary actors that have loose linkages with the inside actors. This kind of structure can be analyzed by the core/periphery structure model (Borgatti & Everett, 1999) in UCINET 6 software (Borgatti et al., 2002).

Core/periphery structure analysis has two kinds of models. One is the discrete model in which the C/P model consists of two classes of nodes: a cohesive subgraph (core) and a class of actors more loosely connected to the core. The other is the continuous model in which each node is assigned a coreness to reflect the extent to which the node is distant from the centroid of a single point cloud in a Euclidean representation. Table 1 and table 2 are the results of the four symmetrized networks processed by using two models of C/P structure. In the light of the core members recommended by the discrete model, we define the members who are all core members in the four symmetrized networks as the core in the network. In the same way, we define the members on the edges in the four symmetrized networks as the periphery in the group. The rest are called semi-periphery. The core members in the group should be the important ones in all of the four networks, not just in one or two kinds of networks. Some persons may be fond of contacting others or consider most of the group members as his friends, but they are incapable of contributing any valuable help. The classification of core/ periphery members just in terms of one or two kinds of networks may result in the wrong conclusions.

Figure 2 Comparison results among the four kinds of Networks. The percentages labeled in the right side of each line are statistical results that are used to show how actors behaved in such three networks as communication, collaboration and help when they are friends or just acquaintances.

Because of little differences between the results of the above two models, we use the analysis results of the discrete model as our final results. By comparing help network with the results of C/P structure, we find that 65.17% of help comes from core members, 27.53% of help from semi-periphery, and only 7.3% of help from peripheral members. The statistical results are shown in table 4.

Table 2 The Analysis Results of C/P structure in four kinds of networks by using the discrete model

Core	Help Network	3 5 7 8 9 10 14 15 18 20 21 22 23 24 29 30 35 38 43 45 49
	Communication Network	3 5 7 8 9 10 14 15 16 17 18 20 21 22 23 24 26 30 35 38 43 45 48 49
	Collaboration Network	5 7 8 9 10 14 17 21 22 23 24 38 45 49
	Friend Network	5 7 8 9 10 14 15 16 20 21 22 23 24 25 29 30 35 38 43 45 48 49
Periphery	Help Network	1 2 4 6 11 12 13 16 17 19 25 26 27 28 31 32 33 34 36 37 39 40 41 42 44 46 47 48 50
	Communication Network	1 2 4 6 11 12 13 19 25 27 28 29 31 32 33 34 36 37 39 40 41 42 44 46 47 50
	Collaboration Network	1 2 3 4 6 11 12 13 15 18 19 20 25 26 27 28 29 30 31 32 33 34 35 36 37 39 40 41 42 43 44 46 47 48 50
	Friend Network	1 2 3 4 6 11 12 13 17 18 19 26 27 28 31 32 33 34 36 37 39 40 41 42 44 46 47 50

Table 3 The Analysis Results of C/P structure in four kinds of networks by using the continuous model

Core	Help Network	3 5 7 8 9 10 14 15 18 20 21 22 23 24 29 30 35 38 43 45 48 49
	Communication Network	3 5 7 8 9 10 14 15 16 17 18 20 21 22 23 24 26 29 30 35 38 43 44 45 48 49
	Collaboration Network	3 5 7 8 9 10 14 16 17 18 21 22 23 24 26 28 29 38 45 48 49
	Friend Network	3 5 6 7 8 9 10 14 16 17 20 21 22 23 24 25 29 30 35 38 43 44 45 48 49
Periphery	Help Network	1 2 4 6 11 12 13 16 17 19 25 26 27 28 31 32 33 34 36 37 39 40 41 42 44 46 47 50
	Communication Network	1 2 4 6 11 12 13 19 25 27 28 31 32 33 34 36 37 39 40 41 42 46 47 50
	Collaboration Network	1 2 4 6 11 12 13 15 19 20 25 27 30 31 32 33 34 35 36 37 39 40 41 42 43 44 46 47 50
	Friend Network	1 2 4 11 12 13 15 18 19 26 27 28 31 32 33 34 36 37 39 40 41 42 46 47 50

Table 4 which block does help come from in C/P structure

Item	Help comes from strong ties (person-time)			Help comes from weak ties (person-time)			Which block does help come from (person-time)		
Blocks	Core	Semi-Periphery	Periphery	Core	Semi-Periphery	Periphery	Core	Semi-Periphery	Periphery
Help	142	56	15	90	42	11	232	98	26
Percentage	66.67	26.29	7.04	62.94	29.37	7.69	65.17	27.53	7.30
Sum	213（60%）			143（40%）			356 (100%)		

DISCUSSION

The network of the academic group as a big cluster is composed of three kinds of actors. The most important actors are the teachers who work at the School of Management, including the supervisor, other professors, and associate professors. They are leaders and organizers of all the academic activities, usually of higher level of academic research and with much experience about how to do management research. Keeping in touch with such actors or collaborating with them will make it easier to get valuable help and access to important ideas and data.

The second kind of actors is the graduates who spend all of their time doing research work. They often take part in academic activities inside and outside the group, and sometimes make their presentations in related academic meetings. Different kinds of information and knowledge are needed for the graduates with different grades. For example, the newcomers who joined in the group for just one or two years need information on references and books, and knowledge about how to search for literature. They are eager to find a problem worth studying. The senior graduates need some knowledge about how to write their dissertation, methods and skills used to solve their difficulties, as well as the information about finding a job.

The last kind of actor is the graduates who spend their spare time finishing their dissertation because giving up their job will lead to a great loss of money. These graduates rarely take part in the academic activities inside the group, but have a lot of information on businesses and much practical experience. They usually are the top or middle managers in their companies or governmental agencies and have many friends or acquaintances in business society. Getting in touch with such actors will make it more convenient to get some information on employment and opportunities for empirical investigation.

Group constitution makes interaction ties characteristic of diversification. The intragroup interaction ties cannot be viewed simply as strong ties, but a complex situation with strong and weak ties interweaved.

Compared with attribute information of the group members (attached in Appendix 1), we can find that the network analysis results of C/P structure are consistent with the real situation in the group. Specifically, core members consist of the mentor, graduate student teachers, and full-time graduates who have been enrolled for a long time (4-8 terms) and have kept ahead in their research work. These graduates often play an active part in academic activities, and their papers and other writings are much more active than the others' in the group. The mentor is a core member in all four networks, and ranks first in the advice and research cooperation networks, ninth in the communication network, and fourteenth in the friendship network.

Periphery members include two categories of members: (1) the full-time graduates who have just enrolled and are beginners in their research field, and (2) the doctoral candidates on duty who have seldom presented in academic activity. These periphery members have little linkage with others in all of the four networks.

Semi-periphery members are the ones who are not core members in all kinds of networks. They play a certain role in the four networks, and will be the core of the group in the future. The activities held in the group are mainly supported by core and semi-periphery members, and the contributions from periphery are insignificant.

Among the total help of 356 (person-time) perceived by group members, 213 (person-time) comes from strong ties (friends), 143 (person-time) from weak ties. That is, sixty percent of help activities comes from friends (strong ties) and forty percent from the others (weak ties). The slight advantage cannot be used to adequately support the hypothesis that help mainly comes from strong ties. This case is just an example of group network characteristics in complex situations. The network analysis in this paper is static analysis of intragroup interactions. Future work is to collect the longitudinal data of the group to give a better example of dynamic networks of the academic group.

In Conclusion, the intragroup interaction ties cannot be viewed simply as strong ties, but the complex situation with strong and weak ties interweaved. Both strong and weak ties are nearly the same importance for the activities of transferring resources in the academic group. The strong ties with the core actors in the group can make it easier to get more valuable help than weak ties. However, new information and possible opportunities are more likely to be provided by weak ties than strong ties.

Appendix 1 The Attribute Variables Table of Members in the Academic Group

No.	Research Interest	Identities	Semesters in the group	Full time	Academic activity	Remark
1	Others	Doctor candidate	Short time	Yes	Often	Study in Japan
2	Others	Doctor candidate	Middle	No	Not often	In the group
3	HeXie Theory	Master candidate	Middle	Yes	Often	Study in Singapore
4	Others	Doctor candidate	Middle	No	Often	In the group
5	Group Decision-making	Doctor candidate	Middle	Yes	Often	Graduated
6	HeXie Theory	Doctor candidate	Short time	Yes	Often	In the group
7	HeXie Theory	Master candidate	Middle	Yes	Often	Doctor candidate
8	HeXie Theory	Master candidate	Middle	Yes	Often	Graduated
9	CMOT	Professor	Very long	Yes	Often	Professor
10	CMOT	Doctor candidate	Short time	Yes	Often	In the group
11	Others	Master candidate	Middle	No	Almost not	In the group
12	Others	Doctor candidate	Middle	No	Almost not	Graduated
13	Others	Doctor candidate	Long	No	Almost not	Graduated
14	Others	Doctor candidate	Long	No	Almost not	Graduated
15	HeXie Theory	Doctor candidate	Very long	No	Often	Associate Prof.
16	HeXie Theory	Doctor candidate	Short time	Yes	Often	In the group
17	Group Decision-making	Doctor candidate	Very long	Yes	Often	Associate Prof.
18	Others	Doctor candidate	Long	Yes	Not often	Associate Prof.
19	HeXie Theory	Doctor candidate	Short time	Yes	Often	In the group
20	HeXie Theory	Master candidate	Short time	Yes	Often	In the group
21	CMOT	Doctor candidate	Long	Yes	Not often	Postdoctoral study
22	HeXie Theory	Professor	Very long	Yes	Often	Professor
23	Organizational Strategy	Associate Prof.	Very long	Yes	Often	Associate Prof.
24	HeXie Theory	Doctor candidate	Long	Yes	Often	Graduated
25	Others	Doctor candidate	Long	No	Not often	Graduated
26	HeXie Theory	Doctor candidate	Short time	Yes	Often	In the group
27	Organizational Strategy	Postdoctoral study	Middle	No	Not often	Graduated
28	CMOT	Postdoctoral study	Short time	Yes	Often	Graduated
29	HeXie Theory	Doctor candidate	Long	No	Often	Assistant Prof.
30	Group Decision-making	Master candidate	Short time	Yes	Often	In the group
31	Others	Doctor candidate	Long	No	Almost not	In the group
32	Others	Doctor candidate	Long	No	Almost not	Graduated
33	Organizational Strategy	Doctor candidate	Long	No	Not often	Graduated
34	Others	Doctor candidate	Short time	Yes	Often	In the group
35	HeXie Theory	Doctor candidate	Middle	Yes	Often	In the group
36	Others	Doctor candidate	Long	No	Not often	In the group
37	Others	Doctor candidate	Middle	No	Almost not	In the group
38	HeXie Theory	Master candidate	Middle	Yes	Often	Graduated
39	Others	Doctor candidate	Middle	No	Almost not	In the group
40	Others	Doctor candidate	Long	No	Almost not	In the group
41	Group Decision-making	Doctor candidate	Short time	Yes	Often	In the group
42	Others	Doctor candidate	Short time	No	Almost not	In the group
43	HeXie Theory	Master candidate	Middle	Yes	Often	Graduated
44	HeXie Theory	Master candidate	Short time	Yes	Often	Graduated
45	HeXie Theory	Doctor candidate	Middle	Yes	Often	Graduated
46	Others	Doctor candidate	Short time	No	Almost not	In the group
47	Others	Doctor candidate	Short time	No	Almost not	In the group
48	HeXie Theory	Master candidate	Middle	Yes	Often	Graduated
49	CMOT	Doctor candidate	Middle	Yes	Often	In the group
50	Others	Doctor candidate	Short time	No	Often	In the group

References

Barsky, N.P. (1999). A Core/Periphery Structure in a Corporate Budgeting Process. *Connections*, 22: 22-29.

Bian, Y. (1997). Bringing Strong Ties Back In: Indirect Connection, Bridge, and Job Searches in China. *American Sociological Review*, 62: 266-285

Bian, Y., Ang S. (1997). Guanxi Networks and Job Mobility in China and Singapore. *Social Forces*, 75: 981-1006

Borgatti, S.P., M.G. Everett. (1999). Models of Core/Periphery Structures. *Social Networks* 21: 375 -395.

Borgatti, S.P., Everett, M.G. and Freeman, L.C. (2002). *Ucinet 6 for Windows: Software for Social Network Analysis*. Harvard: Analytic Technologies.

Brass, D.J., Burkhardt, M.E. (1992). Centrality and Power in Organizations. In N. Nohria and R.G. Eccles (Eds.) *Networks and Organizations: Structure, Form and Action*. Cambridge, MA: Harvard Business School Press, 191-215.

Constant, D., Sproull, L., and Kiesler, S. (1996). The Kindness of Strangers: The Usefulness of Electronic Weak Ties for Technical Advice. *Organization Science*, 7:119-135.

Flache, A. (2002). The Rational Weakness of Strong Ties: Failure of Group Solidarity in a Highly Cohesive Group of Rational Agents. *Journal of Mathematical Sociology*, 26: 189-216.

Granovetter, M. (1973). The Strength of Weak Ties. *American Journal of Sociology*, 78: 1360 - 1380.

Granovetter, M. (1982). The strength of weak ties: A network theory revisited. In P.V. Marsden & N. Lin (Eds.), *Social structure and network analysis*: 105 -131. Beverly Hills: Sage Publications.

Hansen, M.T. (1999). The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge Across Organization Subunits. *Administrative Science Quarterly*, 44: 82-111.

Levin, D.Z., Cross, R. and Abrams, L.C., (2002). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Academy of Management Proceedings*, MOC: D1-D6.

Keister, L.A. (1999). Where do Strong Ties Come From? A Dyad Analysis of the Strength of Inter-firm Exchange Relations during China's Economic Transition. *The International Journal of Organization Analysis*, 7: 5-24.

Kiecher, P., Hartman, C.L. (1994). Predicting Buyer's Selection of Interpersonal Sources: The Role of Strong Ties and Weak Ties. *Advances in Consumer Research*, 21: 464-469.

Krackhardt, D. (1987). Cognitive Social Structures. *Social Networks*, 9:109-134.

Krackhardt, D. (1992). "The Strength of Strong Ties: The Importance of Philos in Organization. In N. Nohria and R.G. Eccles (Eds.) *Networks and Organizations: Structure, Form and Action*. Cambridge, MA: Harvard Business School Press, 216- 239.

Krackhardt, D., Jeffrey, H. (1993). Informal Networks: The Company Behind the Chart. *Harvard Business Review*, 71(4), 104-111.

McClave, J.T. and Terry, S. (2000). *Statistics* (8th ed.). Prentice Hall Inc.(Upper River, New Jersey 07458), Appendix A, table I Random Numbers, p797-799.

Montgomery, J.D. (1994). Weak ties, Employment, and Inequality: An Equilibrium Analysis. *American Journal of Sociology*, 99:1212- 1236.

Nelson, R.E. (1989). The strength of strong ties: social networks and Inter-group conflict in organizations. *Academy of Management Journal*, 32: 377 -401.

Pickering, J.M., King, J.L. (1995). Hardwiring weak ties: inter-organizational computer-mediated communication, occupational communities, and organizational change. *Organization Science*, 6: 479 -486.

Zhigang, T. and Zhu, T. (2000). Agency and Self-Enforcing Contacts. *Journal of Comparative Economics*, 28: 80-94.