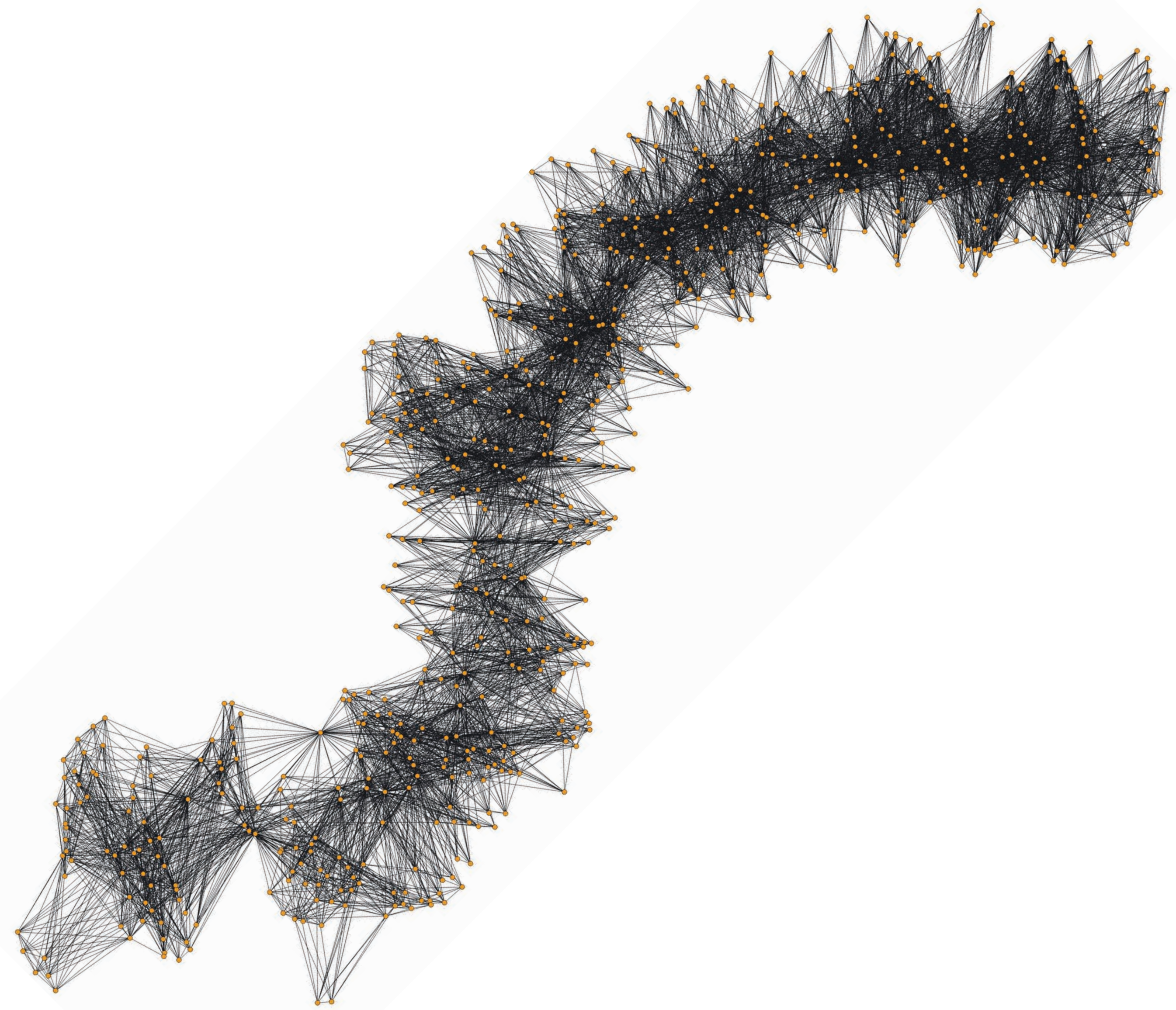
2009 CONNECTIONS Volume 29 • Issue 1

CONNECTIONS

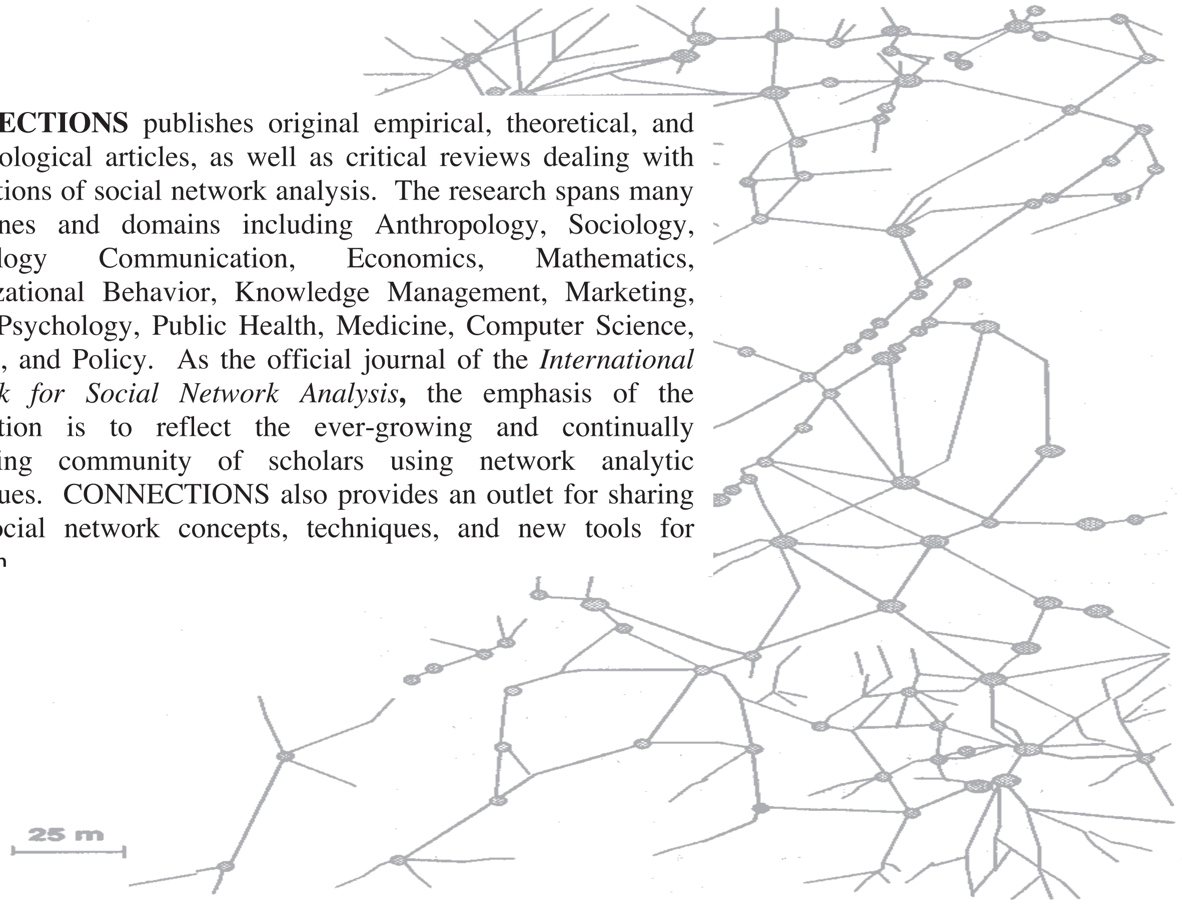
April 2009 *Volume 29 • Issue 1*



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, Mathematics, Organizational Behavior, Knowledge Management, Marketing, Social Psychology, Public Health, Medicine, Computer Science, Physics, and Policy. As the official journal of the *International Network for Social Network Analysis***,** the emphasis of the publication is to reflect the ever-growing and continually expanding community of scholars using network analytic techniques. CONNECTIONS also provides an outlet for sharing new social network concepts, techniques, and new tools for research.



Front Cover: *Images are from enclosed article titled "The Dutch Soccer Team as a Social Network" by Robert Kooij, Almerima Jamakovic, Frank van Kesteren, Tim de Koning, Ildiko Theisler and Pim Veldhoven. Visualization of the Dutch Soccer Team network, in which every node corresponds to a player that has played an official match for the Dutch Soccer Team. A node is connected with another node if both players have appeared in the same match. The network is a small world network, because the average distance between players is small (4.5), while the clustering coefficient is high (0.75).*

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

INSNA includes official board members and five committees:

Board Members

Committees

<u>Finance Committee</u> **-** Chaired by Treasurer Tom Valente

<u>Program/Conference Committee</u> **-** Co-chaired by the host(s) Rebecca Davis, Laura Koehly, and Thomas Valente

<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 complimentary 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

April, 2009 Volume 29 • Issue 1

CONTENTS

The Dutch Soccer Team as a Social Network

Robert Kooij[1,2],
Almerima Jamakovic[2], Frank van Kesteren[2], Tim de Koning[2],
Ildiko Theisler[2], Pim Veldhoven[2]

[1] *Delft University of Technology,*
Faculty of Electrical Engineering,
Mathematics and Computer Science

[2] *TNO Information and Communication Technology,*
Delft, The Netherlands

Abstract

Although being very popular all around the globe, soccer has not received much attention from the scientific community. In this paper we will study the Dutch Soccer Team from the perspective of complex networks. In the DST network every node corresponds to a player that has played an official match for the Dutch Soccer Team. A node is connected with another node if both players have appeared in the same match. The aim of this paper is to study the topological properties of the Dutch Soccer Team network. The motivation for studying the DST network is twofold. The first reason is the immense popularity of the DST, in the Netherlands. Through our study we obtain all kind of new statistics about the DST. Secondly, our results could also be used by the coach of the DST, for instance by determining the optimal line-up. Using data available from a public website we have computed the topological metrics for the DST. Furthermore, we have looked at the evolution of the topological metrics over time and we compared them with those of other real-life networks and of generic network models. We found that the DST is a small world network and that the player with the highest degree also has the lowest clustering coefficient.

Acknowledgments: This research was supported by the Netherlands Organization for Scientific Research (NWO) under project number 643.000.503, and by the Next Generation Infrastructures programme (www.nginfra.nl), which is partially funded by the Dutch government. The authors thank Jos Weber (Delft University of Technology and Excelsior'20) for his valuable comments and suggestions.

Please direct correspondence to Robert Kooij at robert.kooij@tno.nl

Introduction

Soccer is a very popular sport in many countries. According to the coach of the successful AC Milan in the 1990's, Arrigo Sacchi, "it is the most important of the unimportant things in life." Bill Shankly, legendary former manager of Liverpool, made an even more pronounced statement: "Some people say soccer is a matter of life and death. But it is more important than that!"

The popularity of soccer is also reflected in some numbers related to the 2006 World Cup held in Germany. This tournament attracted a cumulative television audience of 27 billion viewers. The global TV coverage was over 73,000 hours (FIFA, 2006).

Although being very popular all around the globe, soccer has not received much attention from the scientific community. For instance, the World Congress on Science & Football, is held only once every 4 years (WCSF, 2007). The edition of this Congress, held in 2007, only attracted 477 attendees, which is not considered a high number for an important scientific Congress.

A nice overview of scientific aspects of soccer is given by Ken Bray (Bray, 2006). In this book, and related references, the following subjects are typically dealt with: physics of the ball, training schemes, performance statistics, medical and physiological aspects, penalty shoot-outs, and the role of electronic devices. Another interesting book, albeit from a completely different perspective, was written by David Winner. In Brilliant Orange, he explores the relation between the Dutch, their history and architecture, their culture and politics, and the influence of each on Dutch Soccer (Winner, 2001).

In this paper, we will study the Dutch Soccer Team from the perspective of complex networks. Our study is inspired by a paper by Onody and De Castro from 2004, who studied a network comprised of Brazilian soccer players (Onody et al., 2004). In this paper, we will study the Dutch Soccer Team (DST) as a social network. In the DST network every node corresponds to a player that has played an official match for the Dutch Soccer Team. A node is connected with another node if both players have appeared in the same match. The aim of this paper is to study the topological properties of the Dutch Soccer Team network.

Studying the topology of real-life networks is important for two reasons. First, it helps us to understand the structure of networks that occur in real-life. Secondly, it can help us to predict how processes on networks evolve. Examples of the latter point include the efficiency of Internet search engines and the spread of viruses on computer networks.

The motivation for studying in particular the DST network is also twofold. The first reason is the immense popularity of the DST in the Netherlands. In particular, Dutch people are very interested in all kinds of facts and statistics related to the DST. Through our study, we obtain many new statistics about the DST. An example of this is "which player had the most co-players?" Secondly, our results could also be used by the coach of the DST, for instance, by determining a line-up where certain aspects of the team are optimal. For instance, a team could be organized so that as many players as possible who have already played together can be on the same team.

The paper is organized as follows. Section 2 describes the topological metrics that will be considered throughout this paper. In Section 3, we describe how we obtained the data and give a visual impression of the DST network. Section 4 discusses results on the topological metrics for the DST network. In Section 5, we give non-network related results that were obtained form the data. Section 6 summarizes our main results and gives some suggestions for further research.

Background

In this section, we provide a set of topological metrics, which is considered relevant in the networking literature (Newman, 2002[a]). A graph theoretic approach is used to model the topology of a complex system as a network with a collection of nodes V and a collection of links E that connect pairs of nodes. A network is represented as an undirected graph *G(V;E)* with $N = |V|$ nodes and $L = |E|$ links.

Link density

The link density S is the ratio of the number of links and the total number of possible links, given by:

$$S = \frac{2L}{N(N-1)} .$$

Degree

The degree d_i of a node i denotes the number of neighbours a node has. The average degree can be easily obtained from the total number of nodes and links:

$E[d_i] = 2L/N$.

Assortativity coefficient

A metric that quantifies the correlation between pairs of nodes is the assortativity coefficient r ($-1 < r < 1$). Networks with $r < 0$ are disassortative, which means that the nodes connect to other nodes with various degrees. In networks with $r > 0$ (assortative networks) the nodes are more likely to connect to nodes with similar degree (Newman, 2002[b]).

The assortativity coefficient r is given by:

$$r = \frac{L^{-1}\sum_i j_i k_i - \left(L^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right)^2}{L^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left(L^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right)^2}$$

,

where j_i and k_i are the degrees of the nodes at the ends of the i-th link, with $i = 1 \ldots L$.

Distance

The distance between a pair of nodes i and j is the length of the shortest path between the nodes. The average distance is the distance averaged over all pairs of nodes.

Diameter

The diameter is the largest distance between any pair of nodes.

Eccentricity

The eccentricity of a node is the largest distance to any other node in the graph. The eccentricity of the graph is the average of eccentricities of all nodes.

Clustering coefficient

The clustering coefficient C_i for a node i is the proportion of links between the nodes within its neighbourhood divided by the number of edges that could possibly exist between the nodes. The clustering coefficient for the whole network is the average of the clustering coefficient for each node.

Closeness

The closeness of a node is the average distance to the other nodes in the graph. Note that some define closeness to be the reciprocal of this quantity. Closeness can be regarded as a measure of how long it will take information to spread from a given node to other reachable nodes in the network. The closeness of a node is a measure of centrality. The node with the lowest closeness is called the most central node.

Algebraic connectivity

The Laplacian matrix of a graph G with N nodes is an NxN matrix $Q = \Delta - A$, where $\Delta = \mathrm{diag}(d_i)$, d_i is the degree of node i and A is the adjacency matrix of G. The second smallest eigenvalue of the Laplacian matrix is called the algebraic connectivity. The algebraic connectivity plays a special role in many problems related to graph theory (e.g. Chung, 1997). The most important is its application to the overall connectivity of a graph: the larger the algebraic connectivity, the more difficult it is to cut a graph into independent components.

Dutch Soccer Team network

The data used to construct the DST network are available at www.voetbalstats.nl. This site contains information about all official soccer matches by the Dutch Soccer Team and about all European matches played by Dutch league teams. A screen shot of this site, which is only available in the Dutch language, is given in Figure 1.

Figure 1: Screen shot of www.voetbalstats.nl

The site gives the line-ups for all official DST matches. We have considered all matches up until Russia - The Netherlands (21 June 2008), which was match number 670. The first match ever of the DST was Belgium – The Netherlands (30 April 1905). As an example we show the line-up of match number 331, The Netherlands – Belgium (18 November 1973), in Figure 2.

Nr. 331

Nederland	0 - 0	België
WKKW	18-11-1973	Toeschouwers: 62000
Stadion: Olympisch Stadion	Bondscoach: Frantisek Fadrhonc	Scheidsrechter: Pavel Khazakov

Speler	Club	Gesc.	Strafs.	In	Uit	e.d.	Int.	Doelp.
Schrijvers Piet	FC Twente						4	
Suurbier Wim	Ajax						25	3
Hulshoff Barry	Ajax						14	6
Mansveld Aad	FC Den Haag						6	
Krol Ruud	Ajax						17	
Haan Arie	Ajax						8	1
Neeskens Johan	Ajax						14	6
Mühren Gerrie	Ajax						10	
Rep Johnny	Ajax						3	1
Cruijff Johan (c)	Barcelona						26	22
Rensenbrink Rob	Anderlecht						10	

Figure 2: Line-up of Match Number 331

All players in Figure 2 appear as nodes in the DST network and are all mutually connected. So, as an example, Aad Mansveld is connected to Rob Rensenbrink.

By working our way through all 670 matches of the DST until June 2008, we have been able to construct the adjacency list of the DST network. In 670 matches, a total number of 691 players appeared. Every individual player was given an ID from 1 to 691. The ID ranking was based on the number of matches played. The adjacency list is the representation of all links in the network as a list. For instance, because Aad Mansveld has ID 294 while Rob Rensenbrink has ID 41, the adjacency list of the DST network contains the entry 41 – 294. We have found that the total number of links in the DST network equals 10,450.

In Figure 3, we have visualized the DST network by importing the adjacency list to the Pajek program (Pajek, 2007).

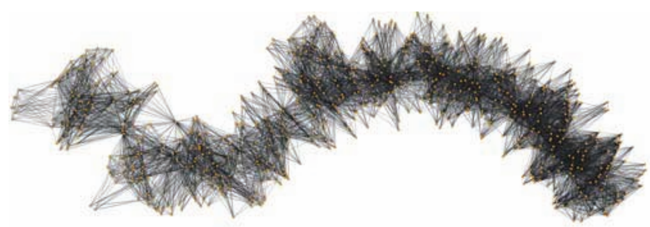


Figure 3: Visualization of the DST network

Nodes on the far left of the graph denote players that played in the beginning of the previous century. Nodes on the far right represent players that were playing in recent years or are still active.

Results

In this section we present the values of the topological metrics introduced in Section 2 for the DST network. We have computed the metrics using Pajek and dedicated Matlab functions. The results are given in Table 1.

Metric	Value
Number of nodes	691
Number of links	10450
Density	0.044
Average degree	30.25
Assortativity coefficient	-0.063
Average distance	4.49
Diameter	11
Eccentricity	8.59
Clustering coefficient	0.75
Algebraic connectivity	0.16

Table 1: Topological Metrics for DST Network

First, we conclude that the DST network is connected, i.e. between every pair of players a path exists. As an example, consider Johan Cruijff and Marco van Basten. These two players never played in the same game. However, they have both played with Willy van de Kerkhof, hence the distance between Cruijff and van Basten is 2.

Because the average distance between players is small (4.49), the DST network exhibits, like many other social networks, the small world phenomenon. In addition, because the clustering coefficient is high (0.75) the DST network is a small world network (Watts, 1999). More detailed information is given in Table 1.

For instance, we are now capable of answering the question “which player had the most co-players?” It turns out that the player with the highest degree is Harry Dénis. In fact, Dénis occurred in matches with 117 other players. On the other hand, Edwin van der Sar, who played the most matches of all players, only has a degree of 97. Interestingly enough, Dénis also has the lowest clustering coefficient (0.17) of all players.

According to Table 1, the diameter of the DST network is 11. For instance, the shortest path between Rafael van der Vaart (who is still an active player) and Jan van Beek has length 11.

Note that the DST network has many other shortest paths of length 11. For instance, any player that only played with Edwin van der Sar after 2000, has a shortest path of length 11 to Jan van Beek. In fact, by using Pajek, we have found that the DST network has 324 shortest paths of length 11.

By computing the closeness of all players, we have been able to determine the most central player in the DST network.

Table 2 shows the top 5 of players with the lowest closeness. The most central player in the DST network is Roel Wiersma, who was active from 1954 to 1962 and played 53 matches. Note that it is not surprising that the most central players were active about 50 years ago, because the Dutch Soccer Team has a history of about 100 years. Of the players still active today, Edgar Davids is most central, with an average distance to the other players of 4.73.

	Player	DST career	Closeness
1	Roel Wiersma	1954-1962	3.119
2	Faas Wilkes	1946-1961	3.213
3	Bertus de Harder	1938-1955	3.217
4	Kees Rijvers	1946-1960	3.222
5	Mick Clavan	1948-1965	3.230

Table 2: Top 5 Most Central Players (lowest closeness)

The evolution of the topological metrics for the DST network over time is given in Table 3.

Metric	1926	1946	1966	1986	2008
Number of nodes	181	282	427	556	691
Number of links	1956	3170	5190	7575	10450
Density	0.12	0.080	0.057	0.049	0.044
Average degree	21.61	22.48	24.31	27.25	30.25
Assortativity coefficient	- 0.17	-0.16	-0.16	-0.11	-0.063
Average distance	2.32	2.70	3.37	3.88	4.49
Diameter	4	6	8	10	11
Eccentricity	3.48	4.58	6.16	7.54	8.59
Clustering coefficient	0.77	0.76	0.76	0.75	0.75
Algebraic connectivity	1.13	0.68	0.31	0.21	0.16

Table 3: Evolution of the DST Network in Time

For the DST network, the number of nodes and links increase over time. It can be observed from Table 3 that also most of the other topological metrics for the DST network increase over time. In fact, the average degree, the average distance, and the diameter all exhibit an almost linear increase in time. Looking at the assortativity coefficient, we conclude that the DST network becomes less disassortative in time. The link density is decreasing in a nonlinear fashion, while the clustering coefficient remains almost constant.

Next, we will compare the topological metrics for the DST network with other real-life networks from nature and society, i.e. technological, social, biological and linguistic networks. For this comparison, which was also reported in Jamakovic et al. (2007), we have considered the following real-life networks:

- Brazilian Soccer Players network (BSP) (Onody et al., 2004)
- the western states power grid of the US (Pow) (Watts et al, 1998)
- American air transportation network (Air) (Colizza et al., 2007)
- the Internet at the autonomous system level (Int) (CAIDA, 2007)
- actors co-appearing in movies (Act) (Barabasi et al., 1999)
- network representing frequent associations between dolphins (Dol) (Lusseau et al., 2003)
- network representing protein interaction of the yeast Saccharomyces cerevisae (Pro) (Jeong et al., 2001)
- network representing word adjacencies in Spanish (Spa) (Milo et al., 2004)

In Table 4, topological metrics for various real-life networks are shown. Some entries in the BSP column are empty because these metrics were not reported in Onody et al. (2004).

Metric	DST	BSP	Pow	Air	Int	Act	Dol	Pro	Spa
Number of nodes	691	13411	4940	2179	20906	10143	62	4713	11558
Number of links	10450	315566	6594	31326	42994	147907	159	19528	43050
Density	0.044	0.0035	0.00054	0.013	0.0002	0.0029	0.084	0.0018	0.00064
Average degree	30.25	47.10	2.67	28.75	4.11	29.16	5.10	8.29	7.45
Assortativity coeff.	-0.063	0.12	0.0036	-0.046	-0.20	0.026	-0.044	-0.13	-0.28
Average distance	4.49	3.29	18.54	3.03	3.89	3.71	3.40	3.16	2.92
Diameter	11	-	46	8	11	13	8	4	10
Eccentricity	8.59	-	34.06	5.87	8.03	9.57	6.50	3.99	7.59
Clustering coeff.	0.75	0.79	0.080	0.48	0.21	0.76	0.26	0.11	0.38
Algebraic conn.	0.16	-	0.0009	0.21	0.015	0.0004	0.17	0.12	0.078

Table 4: Topological Metrics for Various Real-Life Networks

Many observations can be made from Table 4. Here we confine ourselves to just a few. The average degree of the DST network (30.25) has the same order of magnitude as that of the air transportation network and the actor network. On average, a player in the Brasilian league, played with 50% more players, than a player in the DST. The reason for this is probably that far more games are played in a league competition than in a national team. Apart from the power grid network, the DST network has the highest average distance (4.49) between nodes. The clustering coefficients of the DST, the BSP, and the actor networks are comparable and much higher than the other networks considered in Table 4. Next, we compare the topological metrics for the DST network with those of generic network models, such as the random graph of Erdös-Rényi (ER), the small-world graph of Watts-Strogatz (WS), and the scale-free graph of Barabási-Albert (BA) (Bollobás, 2001; Watts, 1999; Barabasi, 2002).

The ER graph is the most investigated topology model (Bollobás, 2001). The most frequently occurring realization of this model is $G_p(N)$, in which N is the number of nodes and p is the probability that there is a link between any two nodes. The major characteristic of $G_p(N)$ is that the existence of a link is independent from the existence of other links. The total number of links in $G_p(N)$ is on average equal to pL_{max}, where $L_{max} = N(N-1)/2$ is the maximum possible number of links. Hence, the link density $q = L/L_{max}$ equals p.

The WS graph captures the fact that, despite the large size of the topology, in most real-world networks, there is a relatively short path between any two nodes. Initially, the WS graph is built on the ring lattice $C(N, k)$, where each of the N nodes is connected to its first $2k$ neighbors (k on either side). Subsequently, a small world is created by moving, for every node, one end of each link (connected to a clockwise neighbor) to a new location chosen uniformly with rewiring probability p_r, such that no double links or loops are allowed. The number of links L in the WS graph, irrespective of p_r, is always equal to $L = Nk$. Hence, the link density satisfies $q = \frac{2k}{N-1}$.

The BA graph gives rise to a class of graphs with a power-law degree distribution. The BA graph is based on two ingredients: growth and preferential attachment of nodes, which implies that nodes with larger degree are more likely candidates for attachment of new nodes. The BA algorithm starts with a small number m_0 of fully-meshed nodes, followed at every time step by a new node attached to $m \leq m_0$ nodes already present in the system. After t timesteps, this procedure results in a graph with $N = t + m_0$ nodes and $L = m_0(m_0-1)/2 + mt$ links. Hence, the link density is

$$q = \frac{m_0(m_0 - 1) + 2mt}{N(N-1)}.$$

Table 5 compares the topological metrics of the DST network and the three considered network models. The values for the metrics for the generic network models are averaged over 1000 simulation runs. The parameters p, k, m, and m_0 are chosen such that all three network models have link density almost identical to that of the DST network. This means that for the WS graph we set k = 15, while for the BA graph we choose $m = 15$ and $m_0 = 31$.

Metric	DST	ER	WS $p_r = 0.1$	WS $p_r = 0.2$	BA
Number of nodes	691	691	691	691	691
Number of links	10450	10450	10365	10365	10365
Link density	0.044	0.044	0.044	0.044	0.044
Average degree	30.25	30.19	30.00	30.00	30.00
Assort. coeff.	- 0.063	- 0.005	0.38	0.34	- 0.005
Average distance	4.49	2.22	8.12	6.51	2.19
Diameter	11	3.00	18.39	14.75	3.00
Eccentricity	8.59	3.00	15.30	11.88	2.96
Clust. coeff.	0.75	0.044	0.72	0.71	0.13
Algebr. conn.	0.16	13.46	0.15	0.21	12.10

Table 5: Comparing the DST Network with Generic Network Models

The main conclusion from Table 5 is that, with respect to the considered topological metrics, the DST network most resembles a WS graph, with rewiring probability $p_r = 0.2$. The only exception is the assortativity coefficient, which is much higher for the WS graph than for the DST network. Note that both the ER and BA graph have a much smaller diameter and clustering coefficient than the DST network.

Non-Network Results

In this next section we give a number of non-network related results about the Dutch Soccer Team. Figure 4 shows the number of matches played by the DST per year.

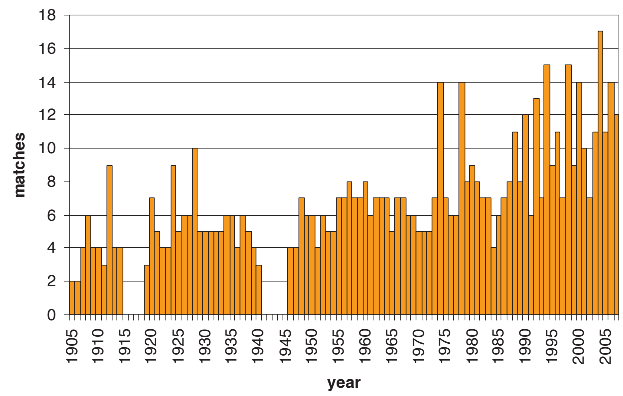


Figure 4: Number of Matches of the DST per Year

A visual inspection of Figure 4 reveals, amongst others, the occurrence of two world wars, local maxima when the DST reached the World Cup final (1974 and 1978), a local minimum when the DST failed to quality for the World Cup (2002) and the trend that the number of matches played per year is increasing.

Because goals are the quintessence of soccer, we will now focus on some goal statistics.

The 10 players that have scored the most goals for the DST are given in Table 6.

	Player	Matches	Goals
1	Patrick Kluivert	79	40
2	Dennis Bergkamp	79	37
3	Faas Wilkes	38	35
4	Abe Lenstra	47	33
5	Johan Cruijff	48	33
6	Ruud van Nistelrooy	64	33
7	Beb Bakhuys	23	28
8	Kick Smit	29	26
9	Marco van Basten	58	24
10	Leen Vente	21	19

Table 1: Leading scorers for DST

Table 1 shows that Patrick Kluivert has scored most goals for the DST. However, we can also see that Kluivert needed more than twice as many games as Faas Wilkes, to score only 5 more goals. For this reason we have also looked at the goal ratio per player, i.e. the number of goals scored by a player per 90 minutes. We only considered players who played 20 matches or more. The result is shown in Table 7.

	Player	Goals	Matches	Minutes	Goals per 90 minutes
1	Beb Bakhuys	28	23	2070	1.22
2	Pierre van Hooijdonk	14	46	1295	0.97
3	Leen Vente	19	21	1870	0.91
4	Faas Wilkes	35	38	3450	0.91
5	Kick Smit	26	29	2587	0.90
6	John Bosman	17	30	1968	0.78
7	Mannes Francken	17	22	2010	0.76
8	Ruud Geels	11	20	1310	0.76
9	Tonny van de Linden	17	24	2138	0.72
10	Abe Lenstra	33	47	4260	0.70

Table 2: Goal Ratio for Players with 20 Matches or More

Of all players that played 20 matches or more, Beb Bakhuys has the highest goal ratio. On this list, Patrick Kluivert is only ranked 14, with a goal ratio of 0.62. It should be noted that Piet de Boer has a goal ratio of 3. He only played once for the DST (match 148 in 1937) and scored three times in this match. The reason that Piet de Boer did not play a second match for the DST is unknown.

Conclusions

In this paper we have studied the topological characteristics of the Dutch Soccer Team network. Taking all matches until June 2008 into account, the main conclusions are as follows:

- The DST network consists of 691 players with 10,450 connections between them.
- The DST network is connected, i.e. between any two players a path exists.
- The DST network is a small world network, because the average distance between players is small (4.49), while the clustering coefficient is high (0.75).
- The player with the most co-players is Harry Dénis, who played together with 117 others.
- Of all players Harry Dénis has the lowest clustering coefficient, i.e. he is the player whose co-players are the least mutually connected.
- The diameter of the DST network is 11, i.e. the longest shortest path has length 11.
- The most central player in the DST network is Roel Wiersma.

Furthermore, we have looked at the evolution of the topological metrics over time. Then, we compared the topological metrics of the DST network with those of other real-life networks and of generic network models.

Finally we have discussed some non-network related results:

- The largest number of matches played by the DST per year is 17. This took place in 2004.
- Of all players that played 20 matches or more, Beb Bakhuys has scored the most goals per 90 minutes (1.22).

Our study reveals a lot of new, interesting statistics, which would best be utilized by the coaches of the DST. The first step towards the development of a decision support tool for coaches would be to examine the topological metrics of players who participated in a particular match and the outcome of the match. The coaches could then determine the line-up for upcoming matches in such a way that certain properties of the team are optimal. For instance, they could choose a line-up so that as many players as possible have already played together. We assume that a team becomes better when enough players have played together before, e.g. because they can anticipate better what the other players are going to do or how they want the ball to be passed to them. Further research could include conducting the same study for the national soccer teams of other countries, the automatic collection and visualization of the DST network, and the development of an interactive tool, which would allow the user to navigate through the DST network, for instance to obtain quickly statistics of a favorite player.

A final possible application of our study is the generation of questions for soccer quizzes.

References

Barabasi, A. L. (2002). Linked: The New Science of Networks, Perseus, Cambridge, MA

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

Bollobás, B. (2001). Random Graphs, Cambridge University Press.

Bray, K. (2006). How to Score: Science and the Beautiful Game, Granta Books.

CAIDA. (2007). Macroscopic Topology AS-Adjacencies Data set, http://www.caida.org/tools/measurement/skitter/as adjacencies.xml

Chung, F. R. K. (1997). Spectral graph theory, Conference Board of the Mathematical Sciences, No.92, AMS, Providence, RI.

Colizza, V., Pastor-Satorras, R., & Vespignani, A. (2007). Reaction-diffusion processes and metapopulation models in heterogeneous networks. Nature Physics, AOP doi:10.1038/nphys560.

FIFA (2006). World Cup and Television, Fifa, http://www.fifa.com/mm/document/fifafacts/ffprojects/ip-401_06e_tv_2658.pdf

Jamakovic, A., Uhlig, S. and Theisler, I.B. (2007). On the relationships between topological metrics in real-world networks, Proc. of the 4th European Conference on Complex Systems (ECCS'07), Dresden, Germany.

Jeong, H., Mason, S., Barabasi, A.L., & Oltvai Z.N. (2001). Centrality and lethality of protein networks. Nature, 411:41.

Lusseau, D., Schneider, K., Boisseau, O.J., Haase, P., Slooten, E., & Dawsen, S.M. (2003). The bottlenose dolphin community of Doubtful Sound features a large proportion of long-lasting associations. Can geographic isolation explain this unique trait?, Behavioral Ecology and Sociobiology, 54: 396-405.

Milo, R., Itzkovic, S., Kashtan, N., Levitt, R., Shen-Orr, S., Ayzenshtat, I., Sheffer, M. and Alon, U. (2004). Superfamilies of evolved and designed networks. Science, 303: 1538-1542.

Newman, M. E. J. (2002[a]). The structure and function of complex networks. SIAM Review, 45: 167-256.

Newman, M. E. J. (2002[b]). Assortative mixing in networks. Phys. Rev. Lett. 89: 208701.

Onody, R.N. and de Castro, P.A. (2004). Complex network study of Brazilian soccer players. Physical Review E, 70 (3): 037103.

Pajek. (2007). Program for large network analysis, http://vlado.fmf.uni-lj.si/pub/networks/pajek/.

Watts, D. J. (2003). Six Degrees: The Science of a Connected Age, W.W. Norton & Company.

Watts, D. J. (1999). Small-Worlds: The Dynamics of Networks Between Order and Randomness, Princeton University Press.

Watts, D. J. and Strogatz S.H. (1998). Collective dynamics of small-world networks. Nature, 393: 440-442.

WCSF. (2007). World Congress on Science & Football, http://www.wcsf2007.org/.

Winner, D. (2001). Brilliant Orange: The Neurotic Genius of Dutch Football, Bloomsbury Publishing PLC.

The Measurement of Social Networks: A Comparison of Alter-Centered and Relationship-Centered Survey Designs

Lyssa L. Thaden
Thomas Rotolo

Department of Sociology
Washington State University

Abstract

Utilizing two surveys administered to a classroom of college students, this study explores differences in social network measures based on survey instrument design. By administering both a relationship-centered survey and an alter-centered survey, we analyze differences in range, mean numbers of relationships, network centralization, and network density. Nonparametric tests are also used to discern patterns of similarity and difference. We find that measurement differences are often negligible when asking about extremely close relationships like friendship. However, differences often appear when studying "weak tie" types of relationships such as recognition of classmate names or acquaintances.

Acknowledgements: The authors thank the reviewers for their helpful comments on an earlier draft of this manuscript.

Please direct correspondence to Lyssa Thaden, Department of Sociology, Washington State University, Pullman, WA 99164-4020. Email: lyssa@wsu.edu.

Introduction

Many important methodological issues have been raised for social network researchers to consider (e.g., Marsden, 1990). In part, a significant amount of research in this area has been devoted to testing survey questions for response effects. Survey researchers have explored the impact of question phrasing, question ordering, and survey design layout on individual response. As a result of this research, a series of recommendations have been developed to help ensure that researchers are able to elicit responses while reducing non-response and measurement error (Braverman and Slater, 1997; Dillman, 2000). Unfortunately for social network researchers, most of this literature is aimed at researchers who have little interest in examining social network structures. The survey methodology literature tends to follow standard neo-classical economic assumptions that treat individuals as independent, atomistic units, and does not focus attention on the methodological challenges associated with measuring relational variables. As such, this literature does not present a coherent set of guidelines for the collection of social network data. Nevertheless, the structure of a survey may very well influence the quality of network data collected.

Fortunately, a number of studies exploring the impact of questionnaire construction and item writing on the quality of social network data collected have recently emerged. Marin (2004) considered the issue of name-generation versus questions about all alters, and found network-level measures differed substantially. Fu (2005) studied daily contacts in personal networks by utilizing two methods: a diary approach and a single-item survey. Kosovsek and Ferligoj (2005) examine differences in response by multiple methods – whether validity and reliability are affected by type of survey (phone vs. in-person) and by question (by alter vs. by question). Following a similar pattern, Coromina and Coenders (2006) checked layout and design of web surveys, including by question/by alter, graphics vs. text, and varying response labels, to assess network data quality.

Adding to this growing body of literature, our research uses social network data obtained from students in a college classroom to determine if differences arise among several network characteristics when we ask respondents about relationships using an *alter-centered* or a *relationship-centered* structured questionnaire. It is our aim in this research to further frame the discussion of survey design effects on the collection of social network data.

Collecting Network Data

As the popularity of collecting social network data with survey techniques has grown, attention has started to focus on assessing the networks' quality of measurement. Of interest to this study is the source and method of collecting information on social ties. Network researchers must make a number of decisions regarding how to measure the relationships between actors. Network researchers wishing to use survey instruments to collect data immediately face two key decisions. First, does the researcher provide the respondent with a list or roster of actors or allow the respondent to use free recall? Second, does the researcher allow a fixed maximum number of alters or an open-ended number of choices for respondents to make when identifying alters (Wasserman and Faust, 1995, p. 46)?

The primary difficulty in utilizing the roster format remains the potentially challenging task of creating an exhaustive list of all of the actors in a social network. For some research settings, a complete roster can be constructed and applied with relative ease, especially when the number of actors is limited. For instance, researchers can often obtain lists for students in a class from an enrollment roster, union participants as specified by due-paying members, or membership lists for other bounded or relatively stable groups. In such cases, a respondent can be presented with an entire list of people in the network and then the researcher can ask the respondent to identify with whom the respondent shares a particular relationship. However, there are times when rosters are not readily available. In some settings, there are too many actors in the system to create a comprehensive list, while in other situations exhaustive rosters are simply unknown. When the researcher cannot construct a complete roster of actors, the respondent can be asked to list those persons with whom the respondent had a particular relationship. For instance, respondents would be asked to "identify all friends." This name-generator approach is likely to work well with questions about relationships that are salient and accessible to the respondent.

Another critical decision that network researchers face involves the number of actors the respondents are allowed to identify on the survey. This issue becomes very important if using the name-generator or free-recall approach. Often, time and space limitations prevent respondents from generating a complete list of actors with whom a particular relationship is shared. Instead, respondents are asked to name only a small number of people with whom they have a particular relationship. For example, respondents might be asked to identify "your three best friends" or "the five people with whom you discuss important matters." Alternatively, respondents might be provided with an initial question without an upper limit on responses, but subsequent follow-up questions might ask information about only the first *n* people the respondent mentioned. This, too, can be problematic. As Holland and Leinhardt (1973) have suggested, by limiting the number of choices a respondent can make, an inherent selection bias may be present and therefore introducing measurement error.

Many of the design issues discussed above were taken into consideration with the first effort to collect representative social network data from the United States population in the 1985 General Social Survey (GSS). Prior to implementation of the GSS social network module, Burt (1984) conducted an extensive review of the amount of time it would take respondents to answer a series of 15 network-related questions, the number of alters a respondent would be asked to name, and the types of relationships about which respondents would be questioned. On the basis of this study, the social network module in the GSS used a name-generator approach to identify alters, limited responses to five alters, and was estimated to take approximately eleven minutes to administer.

In this process, Burt also questions how information about relationships between pairs of alters should be gathered. Burt distinguishes between a "short-form" and a "long-form" questionnaire:

> The short-form variation frames items in terms of a specific kind of relationship. The respondent is asked to identify people between whom the specified relation exists. Are any of these people married to one another? Who among these people dislike one another? (Burt, 1984, p.320).

In contrast, "The long-form instrument, frames items in terms of a specific pair of alters. The respondent is asked to describe the relationship between a specific pair of people" (Burt 1984, p. 321). Thus, in the short-form instrument, the respondent focuses on the relationship and must recall the names of all individuals meeting the condition of the specified relationship. The instrument typically cycles through a number of different types of relationships. The long-form instrument begins with two actors, and asks the respondent to recall the types of relationships shared by these two actors. Then, the respondent must consider the relationships in terms of all other pairs of alters.

In discussing the strengths and weaknesses of each approach, Burt acknowledges possible differences in completion time, reliability, and bias in choosing one form over another. He suggests that the long-form requires more time to administer and would be more likely to tire the respondent, as they assess all possible combinations of actors.

Burt also claims that measures of social network density would be upwardly biased with the long-form because the respondent is asked about the relationship between two specific actors. As Burt puts it, "Given a set of people named as intimates, cognitive balance implies a bias toward perceiving some kind of relation between each pair of intimates" (1984, p. 321). On the other hand, the short-form items require respondents "to evaluate relations for their relative strength, identifying the strongest and weakest" (1984, p. 321). Burt concludes that the differences between short and long forms will result in the short form items producing "greater variability in the structure of interpersonal environments" (1984, p. 322) because the evaluation of the relative strength of the relationship (long-form) will be more stable than the evaluation of the boundaries between relationships (short-form). While Burt empirically tested other aspects of this module, differences between the short and long-form responses were not explored. Instead the short-form was selected for the GSS, primarily because this form takes less time for respondents to complete.

Burt acknowledges the possibility of measurement error that may occur due to differences in how the survey is created and administered. These differences should be of concern in measuring social relationships and evaluating results. If the character and quality of social network data varies by the type of survey instrument used, questions arise about the usefulness of survey approaches, collecting future network data, and the conclusions of prior studies relying on network data collected with such survey techniques. In the following sections, we report on one such effort to explore survey design effects in the collection of social network data.

Data and Methods

The social networks that will be used in this analysis come from one classroom of college students enrolled in a sociology course. Focusing on students as actors in a network allows us to use a roster format for the collection of network data because all members of the group (the students officially enrolled in the class) are known from enrollment records. The classroom had 42 enrolled students. While attendance in the class was quite high, not all students enrolled in the class at the start of the semester were present in the classroom when the surveys were administered. Students not completing the survey were not included as respondents nor were these students counted as alters if named.

The students enrolled in the class do not constitute a random selection from the population of all students. Nevertheless, the composition of students in this class parallels that observed in other offerings of this course. Students in this classroom were homogenous with respect to major, race, and age. No "high-profile" students were enrolled in the class (e.g., male varsity athletes, student government leaders, or other students who might be expected to have larger social networks). While samples drawn from different populations may exhibit different network structures, we have no reason to suspect that our participants would react uniquely to differences in question format.

To assess the possibility of different responses based on question wording, two surveys were constructed, each asking about relationships in different ways. Appendix A provides examples of the layout for each survey. The first survey, the *alter-centered survey*, generates network data by providing respondents with a series of questions about each of his or her classmates and the particular relationships that he or she has with each of these individuals. In the alter-centered survey, the respondent is first given the name of a classmate (alter) and then provided with a list of all of the possible types of relationships that the respondent could have with that classmate. For example, the respondent is given the name "John Smith," followed by a list of possible relationships (recognize his/her name, acquaintance, friend, etc.). The survey continues through a list of each member of the class with each of the possible relationships available for checking.

The second survey, the *relationship-centered survey*, asks respondents a series of questions about the types of relationships the actor has with others in the classroom. First, the respondent is provided with the statement that defines a specific type of relationship. For instance, "I recognize the following person(s) by name…". This phrase is followed by an alphabetized list of all students in the class. Respondents are asked to check all students whom she or he "recognizes by name". The survey continues asking questions in this manner, first identifying a relationship (acquaintance, friend, etc.) and then asking the respondent to mark those classmates that he or she considers having that type of relationship.

The alter-centered instrument resembles Burt's long-form in that respondents must first consider pairs of actors (i.e., the respondent and alter) and then provide information on the relationship. The relationship-centered survey parallels Burt's short-form because respondents are asked to consider a specific type of relationship and then must identify respondent-alter pairs that meet the conditions of the relationship. We expect the alter-centered instrument to be more likely to tire respondents as they shift from one classmate to the next, identifying all relationships that apply, until the entire roster of students is exhausted. As a respondent tires, he or she might not check all relationships that apply for each classmate. Instead, in an effort to complete the survey in an efficient manner, or perhaps because boredom sets in as the respondent considers every student in the class, respondents might resort to checking only the most salient relationships for each respondent. Therefore, a measure of a network characteristic like network degree will likely be underreported in the alter-centered survey. However, we might find, similar to Burt's long form, inflated measures of social network density because the actors are the focus of the survey. The relationship-centered survey will likely produce more variability in the network structure as respondents may have more difficulty distinguishing the boundary of a specific relationship (e.g. acquaintance) when evaluating on a person-to-person basis. That is, difficulty may arise when a respondent must decide whether Actor 1 truly deserves acquaintance status when compared to the acquaintance shared by, say, Actor 2 and Actor 3.

Both survey forms were administered to the same students in the classroom. The alter-centered survey was administered on a Friday and the relationship-centered survey was administered the following Monday. The analyses within the same classroom allow us to see what differences may be revealed by the same people taking two different versions of the survey. This reduces variation in scores that might result due to differences in individuals. However, the possibility of testing effects exists for a within group design. . That is, subjects may remember taking the first survey and retain information to use on the second survey.

Measures

Using the two surveys, we compare the measures of a number of network characteristics. One of the most basic network properties is degree. The degree of an actor or node is the measure of how many other nodes (or alters) the actor is directly connected to, represented as $d(n_i)$ or the degree of node i . The degree of node i can range from zero (a node has no relationship with any other actor in the network) to n-1 (a node has a relationship with every actor in the network with reciprocal ties excepted). This measure provides information about how "connected" nodes are within the network. A group or network-level measure of degree (mean degree) can be developed by summing the degrees of all nodes in the network and dividing by the size of the network (Wasserman and Faust, 1994, p. 100, equation 4.1).

In the first set of analyses, we compare mean degree and other descriptive group-level statistics as measured by the two types of surveys for three salient types of relations that connect students: name recognition, acquaintances, and friends. In the second set of analyses, we compare two additional group-level network properties, density and centrality, for these same three ties: name recognition, acquaintance and friendship relations.

Density is the ratio of the actual number of relationships between people observed in the network and the total number of relationships that are possible within the network. Following Wasserman and Faust (1994, p. 101), if a network of size *g* contains L relationships, the density of the network is defined as:

$$\Delta = \frac{L}{g\,(g-1)},$$

when L=0 there are no relationships between any actors in the network and density equals a minimum of zero. When all nodes are connected to all other actors, L=*g(g-1)* and density equals a maximum of one.

Various measures of network centrality exist in the literature. Our measure of centralization corresponds to "actor closeness centrality," a measure reflecting the closeness (or distance) of all actors in the network to each of the other nodes (Wasserman and Faust 2004, pp. 184-186). We use a standardized measure of actor closeness centrality, ranging from 0 (an isolated node) to 1 (an actor is adjacent to all other actors in the system). As Wasserman and Faust explain, this measure is based on Sabidussi's index (1966) of actor closeness, "the inverse sum of the distances from actor *i* to all the other actors" (1994, p. 184). In mathematical form, the measure is specified as:

$$C_C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j)\right]^{-1} (g-1),$$

where $d(n_i, n_j)$ indicates the smallest number of lines linking nodes n_i and n_j (the geodesic distance). This distance function is summed over all network actors, *g*, and the closeness centrality measure is calculated by taking the reciprocal of this sum. The standardized actor closeness centrality measure is multiplied by the size of the network minus unity, *(g-1)*, in order to allow for an accurate comparison across networks of varying size.

Finally, we conduct nonparametric tests for differences in a variety of network characteristics across surveys. We use non-parametric tests because the assumption of normality is not met for the network measures we consider. And, while t-tests and other statistical tests based on the arithmetic mean are subject to influence by possible outliers, our non-parametric tests are not. In particular, we conduct a median test to assess whether two or more samples are drawn from populations with the same median.

Results

Respondents were asked about seventeen different relationships they had with other members of the class. In this analysis, we focus on three of the most often cited relationship ties: recognizing the name of other classmates, considering a classmate an acquaintance, and considering a classmate a friend. The degree, density and network centralization of actors is then considered within each of these relationship types across the two surveys administered. All social network measures were calculated using UCINET 6 Social Network Analysis Software (Borgatti, Everett, & Freeman, 2002).

Table 1 presents the results for the group mean degree, standard deviation, median degree, and minimum and maximum values of degree observed. For two of the three relationships observed within the classroom, the relationship-centered survey produces more alters named by respondents as compared to the alter-centered survey. (The third relationship – recognize name – resulted in the same mean value.) In other words, respondents failed to identify some of the existing relationships that were measured in the alter-centered survey. These results provide support for the idea that respondents became tired or perhaps identified only the most salient relationships when responding to the alter-centered survey.

Table 1. Degree Measures

Relationship	Classroom	Mean	S.D.	Median	Minimum	Maximum	N
Recognize Name	Alter-Centered	6.90	8.22	4.5	0	38	42
	Relationship-Centered	6.90	3.62	6	1	18	42
Acquaintance	Alter-Centered	2.10	1.78	2	0	9	42
	Relationship-Centered	3.40	3.15	3	0	21	42
Friend	Alter-Centered	1.57	1.02	2	0	3	42
	Relationship-Centered	2.14	1.24	2	0	4	42

This finding is further supported by considering the total number of relationships identified by respondents in each survey. Table 2 shows that the relationship-centered survey produced an average of 18.33 relationships identified by each respondent, while the alter-centered survey resulted in only 15.26 relationships. Thus, when the same respondents are asked about relationships with classmates using two different survey instruments, the relationship-centered survey identified more relationships than the alter-centered survey.

Table 2. Total Number of Relationships Recognized

Classroom	Mean # of Relationships	S.D.	Median	Min	Max
Alter-Centered	15.26	9.78	12	1	46
Relationship-Centered	18.33	8.92	17	5	49

Table 3 provides another look at these relationships by providing the network density and centralization measures. Here, too, we find for the three major relationships (recognize name, acquaintance, friend), network density is higher in the relationship-centered survey (recognize name density = 0.1678, acquaintance density = 0.0830, friend density = 0.0523) compared to the alter-centered survey (recognize name density = 0.0523, acquaintance density = 0.0511, friend density = 0.0383). While these reported differences are small, this pattern of within classroom difference for density holds when looking at even less salient relationships, such as the identification of homework partners or shared organizational membership. Similarly, we find that the relationship-centered survey produces higher measures of centralization for both name recognition (0.010 compared to 0.001) and acquaintance (0.012 compared to 0.005). This would seem to indicate that, as reported above, the relationship-centered instrument produces networks with higher levels of connectedness between actors in this system.

Table 3: Centralization and Density Measures

Relationship	Classroom	Density	S.D. of Personal Network Densities	Network Centralization	N
Recognize Name	Alter-Centered	0.0523	0.2226	0.001	42
	Relationship-Centered	0.1678	0.3737	0.010	42
Acquaintance	Alter-Centered	0.0511	0.2202	0.005	42
	Relationship-Centered	0.0830	0.2259	0.012	42
Friend	Alter-Centered	0.0383	0.1920	0.001	42
	Relationship-Centered	0.0523	0.2226	0.001	42

Non-parametric Tests

We also test for significant differences between the survey instruments with tests for variations in the median across classrooms and across relationships. We find no statistical difference in median degree for the relationship of "recognize name." However, we do find a significant difference in the median when respondents are asked to identify acquaintances ($p < 0.05$) or friends ($p < 0.05$). This implies that these two classrooms are not from populations with the same medians, although they are in fact the same students. These findings suggest that differences in the structure of the survey instruments produce differences in the respondents' network structures.

Discussion

The results presented above leave us with evidence that differences occur between alter-centered and relationship-centered question formatting. Clear differences emerge when we examine the broadest relationship-type – recognizing a classmate by name. Here, the range and standard deviation of alters named is much greater when the alter-centered survey is used. Similar differences are observed with the network centralization measure; the alter-centered survey once again yields significantly higher estimates than its relationship-centered counterpart. These findings substantiate Burt's (1984) intuition regarding the reliability of respondents' choices due to the focus on one individual classmate at a time.

Contrary to Burt's (1984) hypothesis, however, upwardly biased estimates of network density do not appear when using the alter-centered form of the questionnaire. In fact, the results indicate the opposite – although the difference is not statistically significant. This finding is consistent not just with the recognition of name relationship; similar conclusions are reached when considering acquaintance and friendship ties.

Network degree and centralization are more stable when evaluating the acquaintance and friendship ties. Here, unlike the relationship involving the recognition of a classmate's name, variation is greatly reduced. When we look at the differences between surveys, it appears as though the relationship-centered survey produces greater recognition of acquaintances and friends. However, the range in degree minimum and maximum, along with the standard deviations, are reduced considerably. The same is true with the network centralization measure.

These results lead us to conclude that meaningful differences in measurement are not apparent when we ask about extremely close relationships like friendship ties. In this case, we suggest Burt's (1984) consideration of time is most important. Results of this study did find that respondents completed the relationship-centered survey more quickly (3.91 minutes) than they completed the alter-centered survey (4.67 minutes). This is similar to what Burt found with the short-form versus long-form of the GSS module.

However, when we ask respondents about people they know less well, researchers would do well to consider the type of survey instrument used to elicit response. Clearly, greater connections, as measured by network density and centralization, were acknowledged when the relationship-centered survey was used to elicit recognition of other classmates by name. In a world where weak ties are acknowledged to be of great importance for many outcomes (Granovetter 1973), accurate measurement of these weak ties must be obtained.

Certainly, this examination is only a beginning. Additional research about how social network researchers should best measure relationships must be pursued. Explorations of other populations of actors and relationships would be beneficial. Further, future studies should consider different modes of administration. A test of the differences explored in this study using web-based surveys may yield very interesting results – especially among young college-aged respondents who are familiar with this technology. Utilizing a split-half study in these types of large groups, either by web or paper survey, may provide further evidence that confirms the results found in this study. Additionally, looking at other network measures, such as individual measures of centrality, reciprocal ties between actors, and even clique structure, may help to further clarify differences that emerge based on instrumentation.

References

Borgatti SP, Everett MG and Freeman LC. 2002. *Ucinet for Windows: Software for social network analysis*. Harvard, MA: Analytic Technologies.

Braverman MT and Slater JK. 1997. *Advances in survey research*, San Francisco: Jossey-Bass.

Burt RS. 1984. Network items and the General Social Survey. *Social Networks*, 6:293-339.

Canache D, Mondak JJ, and Seligson, MA. 2001. Meaning and measurement in cross-national research on satisfaction with democracy. *The Public Opinion Quarterly*, 65(4):506-528.

Coromina L and Coenders G. 2006. Reliability and validity of egocentered network data collected via web: A meta-analysis of multilevel multitrait multimethod studies. *Social Networks,* 28:209-231.

Costenbader E and Valente TW. 2003. The stability of centrality measures when networks are sampled. *Social Networks*, 25:283-307.

Dillman DA. 2000. *Mail and internet surveys: The tailored design method*, New York: J. Wiley.

Fu Y. 2005. Measuring personal networks with daily contacts: A single item survey question and the contact diary. *Social Networks*, 27:169-186.

Hamilton, Lawrence. 1996. *Data analysis for social scientists*, Belmont, CA: Wadsworth Publishing Company.

Granovetter M. 1973. The strength of weak ties. *American Journal of Sociology*, 78:1360-1380.

Jolliffe D. 2001. Measuring absolute and relative poverty: The sensitivity of estimated household consumption to survey design. *Journal of Economic and Social Measurement*, 27: 1-23.

Kogovsek T and Ferligoj A. 2005. Effects on reliability and validity of egocentered network measurements. *Social Networks*, 27:205-229.

Lin N, Fu Y, and Hsung R. 2001. The position generator: Measurement techniques for investigations of social capital. In *Social Capital: Theory and Research*, N. Lin, K. Cook and R. Burt (eds.), Hawthorne, New York: Aldine de Gruyter.

Marin A. 2004. Are respondents more likely to list alters with certain characteristics? Implications for name generator data. *Social Networks*, 26:289-307.

Marsden PV. 1990. Network data and measurement. *Annual Review of Sociology*, 16:435-463.

Mayhew BH. and Schollaert PT. 1981. The concentration of wealth: A sociological

model. *Sociological Focus*, 13:1-35.

Sabidussi, G. 1966. The centrality index of a graph. *Psychometrika*, 31:581-603.

StataCorp. 2005. *Stata statistical software: Release 9*. College Station, TX: StataCorp LP.

Wasserman S and Faust K. 1994. *Social network analysis: Methods and applications*. New York: Cambridge University Press.

Appendix A. Alter-centered and Relationship-Centered Surveys

Alter-Centered Survey

Student 1:
□ Recognize his/her name
□ Acquaintance
□ Friend
□ Significant Other

□ Classmate (outside of this course)
□ Homework/study partner
□ Co-worker
□ Neighbor
□ Roommate

□ Attend similar social events
□ Belong to the same organization
(check all that apply)
□ Greek Fraternity/Sorority
□ Intermural Sports Team
□ Religious Organization
□ Residence Hall Association
□ Other ____________________

□ Other relationship __________________

□ Other relationship __________________

Student 2:
□ Recognize his/her name
□ Acquaintance
□ Friend
□ Significant Other

□ Classmate (outside of this course)
□ Homework/study partner
□ Co-worker
□ Neighbor
□ Roommate

□ Attend similar social events
□ Belong to the same organization
(check all that apply)
□ Greek Fraternity/Sorority
□ Intermural Sports Team
□ Religious Organization
□ Residence Hall Association
□ Other ____________________

□ Other relationship __________________

□ Other relationship __________________

Student 3:
□ Recognize his/her name
□ Acquaintance
□ Friend
□ Significant Other

□ Classmate (outside of this course)
□ Homework/study partner
□ Co-worker
□ Neighbor
□ Roommate

□ Attend similar social events
□ Belong to the same organization
(check all that apply)
□ Greek Fraternity/Sorority
□ Intermural Sports Team
□ Religious Organization
□ Residence Hall Association
□ Other ____________________

□ Other relationship __________________

□ Other relationship __________________

Student 4:
□ Recognize his/her name
□ Acquaintance
□ Friend
□ Significant Other

□ Classmate (outside of this course)
□ Homework/study partner
□ Co-worker
□ Neighbor
□ Roommate

□ Attend similar social events
□ Belong to the same organization
(check all that apply)
□ Greek Fraternity/Sorority
□ Intermural Sports Team
□ Religious Organization
□ Residence Hall Association
□ Other ____________________

□ Other relationship __________________

□ Other relationship __________________

Relationship-Centered Survey

1. I recognize the following person(s) by name.

□ Student 1	□ Student 19	□ Student 37
□ Student 2	□ Student 20	□ Student 38
□ Student 3	□ Student 21	□ Student 39
□ Student 4	□ Student 22	□ Student 40
□ Student 5	□ Student 23	□ Student 41
□ Student 6	□ Student 24	□ Student 42
□ Student 7	□ Student 25	□ Student 43
□ Student 8	□ Student 26	□ Student 44
□ Student 9	□ Student 27	□ Student 45
□ Student 10	□ Student 28	□ Student 46
□ Student 11	□ Student 29	□ Student 47
□ Student 12	□ Student 30	□ Student 48
□ Student 13	□ Student 31	□ Student 49
□ Student 14	□ Student 32	□ Student 50
□ Student 15	□ Student 33	□ Student 51
□ Student 16	□ Student 34	□ Student 52
□ Student 17	□ Student 35	□ Student 53
□ Student 18	□ Student 36	□ Student 54

2. I consider the following person(s) to be an acquaintance.

□ Student 1	□ Student 19	□ Student 37
□ Student 2	□ Student 20	□ Student 38
□ Student 3	□ Student 21	□ Student 39
□ Student 4	□ Student 22	□ Student 40
□ Student 5	□ Student 23	□ Student 41
□ Student 6	□ Student 24	□ Student 42
□ Student 7	□ Student 25	□ Student 43
□ Student 8	□ Student 26	□ Student 44
□ Student 9	□ Student 27	□ Student 45
□ Student 10	□ Student 28	□ Student 46
□ Student 11	□ Student 29	□ Student 47
□ Student 12	□ Student 30	□ Student 48
□ Student 13	□ Student 31	□ Student 49
□ Student 14	□ Student 32	□ Student 50
□ Student 15	□ Student 33	□ Student 51
□ Student 16	□ Student 34	□ Student 52
□ Student 17	□ Student 35	□ Student 53
□ Student 18	□ Student 36	□ Student 54

Using SAS to Calculate Betweenness Centrality

Alan R. Ellis

Research Associate and Fellow
Cecil G. Sheps Center for Health Services Research
University of North Carolina at Chapel Hill

Abstract
Betweenness centrality is a useful measure of an actor's importance in a social network. The SAS PROC IML module presented in this paper facilitates the calculation of betweenness centrality by social scientists by making it possible to run a faster algorithm for betweenness centrality using popular statistical software. The algorithm and module could be extended in order to calculate betweenness centrality for weighted graphs or to calculate other network measures that are based on geodesics, such as closeness centrality, graph centrality, or radiality.

Acknowledgments: This research was supported by grant 1-R03-MH073728-01A1 from the National Institute of Mental Health.

Please address correspondence to Alan R. Ellis at 725 MLK Blvd. CB 7590, Chapel Hill, NC 27599-7590. Phone: (919) 966-2340, Fax: (919) 966-1634, Email: are@unc.edu.

Introduction

In analyzing data on social networks, researchers are often interested in *betweenness centrality,* which is one measure of an actor's importance in a network. Betweenness centrality reflects the extent to which an actor lies on geodesics between others in the network. Depending on the context, betweenness centrality might indicate the degree of power, control, or stress experienced by an actor in the course of network interactions (Freeman, 1977). The idea was introduced by Bavelas (1948, cited in Freeman, 1977), and the measure was defined by Freeman (1977):

$$C_B(p_k) = \sum_{i=1}^{j-1}\sum_{j=1}^{n}\frac{g_{ij}(p_k)}{g_{ij}}, i \neq j \neq k$$

(1)

where p_k is a point on the graph, n is the total number of vertices, i and j index vertices on the graph other than p_k, g_{ij} is the number of geodesics between a pair of vertices, and $g_{ij}(p_k)$ is the number of such geodesics that include p_k. C_B, then, is the proportion of geodesics between others in the network on which actor p_k lies. As defined by Freeman, betweenness centrality is normalized by dividing by its maximum possible value, which is the number of vertex pairs excluding p_k:

$$C'_B(p_k) = \frac{C_B(p_k)}{(n-1)(n-2)/2}$$

(2)

The normalized version of betweenness centrality ranges from 0 to 1 and allows comparisons between networks.

There are two significant barriers to the use of these measures of betweenness centrality by social scientists: (1) the need for specialized network analysis software and (2) the space and time typically required for processing. A related measure that requires less computation is egocentric network centrality (i.e., centrality of an actor within its first-order zone), which may predict betweenness centrality fairly well. Marsden (2002) found correlations ranging from .83 to .99 in an analysis of network data from a variety of studies with network sizes from 14 to 217. More recently, Everett and Borgatti (2005) found average correlations ranging from .85 to .99 in simulations with 25-500 actors and network density ranging from .1 to .6. However, a correlation of .83 implies that in the worst case tested, 31% (=1-.83^2) of the variance in betweenness centrality was not explained by the egocentric measure.

Furthermore, Marsden identified scenarios in which the egocentric measure does a poor job of predicting betweenness centrality for a particular actor. (These involved the index actor having alters with extremely high or extremely low centrality in their own first-order zones.) Rather than perform analyses that are subject to this type of error, it is preferable to calculate the standard measure of betweenness centrality if data on the full network are available. This paper diminishes the barriers to calculating betweenness centrality by introducing a SAS Interactive Matrix Language (IML; SAS Institute, 2007) module that implements the faster algorithm for betweenness centrality recently developed by Brandes (2001).

A SAS PROC IML module to calculate betweenness centrality

Brandes' (2001) strategy is to count and store a network's geodesics more quickly using network traversal algorithms instead of matrix multiplications. The author implemented Brandes' algorithm for unweighted graphs using a SAS PROC IML module and supporting macros (Appendix 1; also available from the author's website, http://www.unc.edu/~arellis). The supporting macros use row vectors to implement stacks and queues to store information obtained during network traversal. Stacks are "last-in, first-out" storage mechanisms into which a unit of information can be "pushed" and out of which a unit of information can be "popped". Queues are "first-in, first-out" storage mechanisms into which a unit of information can be "enqueued" and out of which a unit of information can be "dequeued". The PROC IML module expects as input a square matrix (which should be symmetric if the graph is undirected), a dummy (0-1) variable to indicate whether the graph is directed, and another dummy variable to indicate whether betweenness centrality should be normalized. Based on this input, the module traverses the network and returns a column vector which contains, for each vertex in the graph, the requested form of betweenness centrality.

Performance on simulated networks

The module was used to calculate betweenness centrality for 30 simulated networks with size ranging from 100-1000 vertices (100, 200, …, 1000) and with densities of .25, .50, and .75. The module was run on a POWER5+ processor server running AIX UNIX. The maximum amount of memory used was 47 MB. Figure 1 shows the amount of processing time required as a function of network size and density. For the network with 1000 vertices and a density of .75, the processing time was 5,318 seconds (approximately 89 minutes).

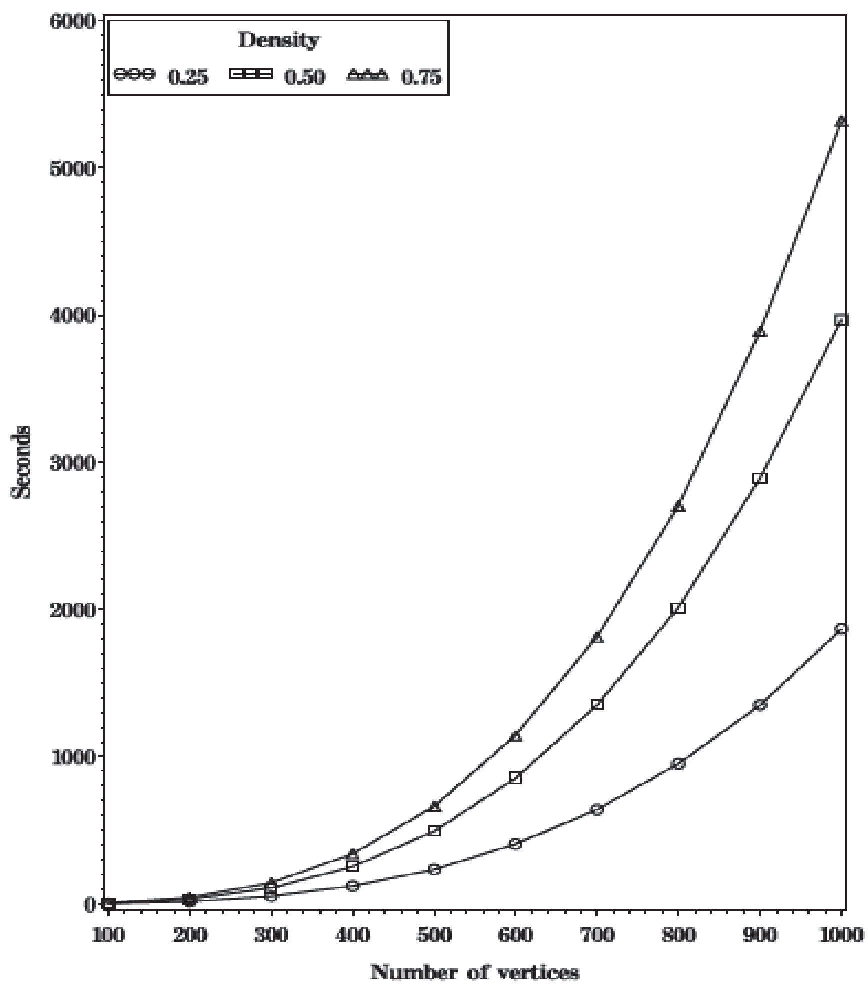


Figure 1. Processing time as a function of network size and density

Given current computer technology, the memory required for running the module is negligible even for a network with 1000 vertices. Therefore, processing time is more important than memory as a potential barrier to data analysis. According to Brandes (2001), processing time should be on the order of $n*m$, where n is network size and m is the number of links in the network, equal to $p*n(n-1)/2$ where p is the network density (Scott, 2000). For the simulated networks, processing time was indeed on the order of $n*m$. This was verified with a linear regression model that predicted processing time as a function of n^3*p *(R^2=.996; other* *results not shown)*. This means that (1) for networks of a given density, the processing time required is roughly proportional to n^3, and (2) for networks of a given size, the processing time required is roughly proportional to p.

Conclusion

Betweenness centrality is a useful measure of an actor's importance in a social network. The SAS PROC IML module presented in this paper facilitates the calculation of betweenness centrality by social scientists by making it possible to run Brandes' (2001) faster algorithm for betweenness centrality using popular statistical software. As noted by Brandes (2001), his algorithm, and therefore the module presented here, could be extended in order to calculate betweenness centrality for weighted graphs or to calculate other network measures that are based on geodesics such as closeness centrality (Sabidussi, 1966), graph centrality (Hage & Harary, 1995), or radiality (Valente & Foreman, 1998).

References

Bavelas, A. (1948). A mathematical model for group structure. *Applied Anthropology*, 7, 16-30.

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*, 25(2): 163-177.

Everett, M. & Borgatti, S. P. (2005). Ego network betweenness. *Social Networks*, 27, 31-38.

Freeman, L.C. (1977). A set of centrality measures based on betweenness. *Sociometry*, 40, 35-41.

Hage, P. & Harary, F. (1995). Eccentricity and centrality in networks. *Social Networks*, 17, 57-63.

Marsden, P. V. (2002). Egocentric and sociocentric measures of network centrality. *Social Networks*, 24, 407-422.

SAS Institute. (2007). SAS OnlineDoc 9.1.3. Retrieved January 30, 2008, from http://support.sas.com/onlinedoc/913/docMainpage.jsp

Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31, 581-603.

Scott, J., (2000). *Social network analysis*. (2nd ed.). Thousand Oaks, CA: SAGE.

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(1), 89-105.

Appendix 1. SAS PROC IML module and supporting macros

```
/*      bcent.sas

        2007/09/12

        [Alan R. Ellis, MSW, http://www.unc.edu/~arellis; emial: are@unc.edu]

        IML modules and supporting SAS macros to calculate betweenness centrality
        Uses algorithm by Ulrik Brandes
            Brandes, Ulrik. (2001). A faster algorithm for betweenness centrality.
            Journal of Mathematical Sociology, 25(2), 163-177.

        returns a vector of betweenness centrality values

        Usage:  %include bcent.sas;
                bc=bcent(m,directed,normalize)
                    m = matrix for which betweenness centrality is to be computed
                    directed = 0 if relationships are undirected, 1 if directed
                    normalize = 0 if raw centrality is desired, 1 if normalized value is desired

        This module creates the following macros: push, pop, enq, deq, isempty

        Input matrix should be square and, if relationships are undirected, should be symmetric.

        2006/04/11   original version
        2006/05/04   changed code so input matrix would not be modified
        2007/08/31   revised comments
        2007/09/12   changed name of input matrix and disabled error checking code that modified input matrix
*/

/* push values onto a stack - i.e., insert into column 1 of a row vector */
%macro push(s,val);
        if ncol(&s)=0 then                      /* if empty then start with value */
                &s=&val;
            else
                &s=insert(&s,&val,0,1);         /* otherwise insert value */
%mend;

/* pop value off a stack - i.e., remove from column 1 of a row vector */
%macro pop(s,val);
        if ncol(&s)=0 then do;                  /* if empty then return undefined value */
            free &val;
            end;
        else do;
            &val=&s[1];
            &s=remove(&s,1);
        end;
%mend;

/* enqueue a value - i.e., insert it at the end of a row vector */
/* adapted from "push" macro - simply changed location of insertion */
%macro enq(s,val);
        if ncol(&s)=0 then                              /* if empty then start with value */
                &s=&val;
            else
                &s=insert(&s,&val,0,1+ncol(&s));   /* otherwise insert value at end of queue */
%mend;

/* dequeue a value - i.e., remove it from column 1 of a row vector */
%macro deq(q,val);
        %pop(&q,&val);
%mend;

/* function to determine whether a stack or queue (i.e., row vector) is empty */
%macro isempty(sq);
        %str((ncol(&sq)=0))
%mend;

/* function to calculate Betweenness Centrality using Brandes' algorithm */
start bcent(m,directed,normalize);
        nvert=nrow(m);                          /* # vertices in network */

        /* check input */
```

```
err=0;
if ((directed^=0) & (directed^=1) & (normalize^=0) & (normalize^=1)) then
     err=1;

/*  The following lines could be enabled in order to check input further.  Note that the input
     matrix is modified. */

/*
else if (nvert^=ncol(m)) then
     err=1;
else do;
     tm=m`;
     if ((directed=0) & any(m^=tm)) then err=1;
end;
*/

if err=1 then do;
     file log;
     put 'ERROR: invalid parameter values for bcent().  Returning -1.';
     put 'USAGE: bcent(m,directed,normalize);';
     put '       <m>        : square matrix (symmetric if graph is undirected)';
     put '       <directed> : 1 if graph is directed, 0 otherwise';
     put '       <normalize>: 1 if result should be normalized, 0 otherwise';
     return(-1);
end;
cb=j(nvert,1,0);                    /* betweenness centrality of each vertex starts at zero */

do s=1 to nvert;
     free stack;                         /* start with empty stack - just being explicit */
     p=j(nvert,nvert,0);                 /* create empty list of predecessors for each vertex */

     sigma=j(nvert,1,0);                 /* count # geodesics each vertex is on: initially zero, */
     sigma[s]=1;                         /* except one for current vertex */

     d=j(nvert,1,-1);               /* vector of -1 values, except zero for current vertex */
     d[s]=0;                             /* d appears to measure depth */

     free queue;                         /* start with empty queue - just being explicit */
     %enq(queue,s);                      /* add current vertex to queue */

     do while (%isempty(queue)^=1); /* while queue is not empty */
          %deq(queue,v);                 /* de-queue a vertex number into v */
          %push(stack,v);                /* and also push it onto the stack */

          /* loop through each neighbor w-sub-j of v */
          w=loc(m[v,]);             /* loop through vertices w where m(v,w) is nonzero */
                                         /* i.e., there is a path from v to w */
          do j=1 to ncol(w);

               /* w-sub-j found for the first time? */
               if d[w[j]]<0 then do;
                    %enq(queue,w[j]);
                    d[w[j]]=d[v]+1;
               end;

               /* shortest path to w via v? */
               if d[w[j]]=d[v]+1 then do;
                    sigma[w[j]]=sigma[w[j]]+sigma[v];
                    p[w[j],v]=1;     /* add v to w's list of vertices */
               end;
          end;
     end;

     delta=j(nvert,1,0);                      /* initialize delta to zero for each vertex */

     /* stack returns vertices in order of non-increasing distance from vertex s */

     do while (%isempty(stack)^=1);     /* while stack is not empty */
          %pop(stack,ww);                     /* use double-w; this is distinct from */
                                              /* the w used for neighbors-to-v above */

          vv=loc(p[ww,]);                     /* indices of vertices on ww's list of predecessors */
                                              /* use vv - this is a new v, too */
```

```
            do k=1 to ncol(vv);
                delta[vv[k]]=delta[vv[k]]+(sigma[vv[k]]/sigma[ww])*(1+delta[ww]);
            end;
            if ww ^= s then cb[ww]=cb[ww]+delta[ww];
        end;
    end;
    if (directed=0) then cb=cb/2;                                /* if undirected then divide by 2 */
    if (normalize=1) then do;
        if (directed=0) then cb=2*cb/(nvert-1)/(nvert-2);       /* normalize by maximum possible centrality */
        else cb=cb/(nvert-1)/(nvert-2);
    end;
    return(cb);
finish;
```

Different States, Choice, Structure and Aggregation in Simulated Social Networks

Rob Stocker

School of Information Technology and Electrical Engineering
&
Defence and Security Applications Research Centre
The Australian Defence Force Academy

Abstract

The fabric of society lies in the networks of connections and patterns of communication that its members create deliberately or inadvertently. Information, ideas, values and norms are passed across this fabric and members can form aggregates or allegiances that centre on common interests, goals, attitudes and the like. Simple multi-agent models of social networks have provided useful insights into the emergence of global network behavior where agents have limited or binary choices of state. This paper examines the impact that a greater number of choices of state have on the emergence of clusters of agents. It examines the global behavior of static populations of interacting computational agents, connected in fixed networks structures that are faced with multiple choices of state. Results indicate that aggregation around a few states appears to be a universal property, independent of network structure.

Acknowledgements

I gratefully acknowledge Professor David G. Green (CSSE Monash University, Melbourne), Dr. David Cornforth (IT&EE UNSW@ADFA, Canberra) Dr. David Newth (CSIRO Canberra) and Associate Professor Russell Standish (UNSW Sydney) for their advice and support in this research.

Please address correspondence to Bob Stocker at r.stocker@adfa.edu.au.

Introduction

How does the process of socialization occur? How are society's values, norms, and ideals cemented in social structure? How do rumours, urban legends and myths spread in society? Why are some people better informed than others? How do people form into groups? What influences the outcome of political elections? These questions create strong interest for social network researchers. Typical of complex adaptive systems, social networks exhibit clustering behavior that may be based around a common principle.

Complex systems are characterised by large numbers of components that interact and communicate through patterns of connections called networks (Holland 1995). Complex systems behavior is at the root of many natural and artificial phenomena. However, knowledge and understanding about their behavior, design and management remains largely empirical. The complexity of multi-agent systems means that traditional methods of studying them are not effective (Bura *et al.* 1995; Gilbert & Troitzsch 1999; Goldspink 2002; Wassermann 1980), as they tend to be unstable and unpredictable (Bak & Sneppen 1993; Erdös & Rènyi 1960; Green 1993; Horgan 1995; Langton 1990).

Structural analysis (Berkowitz 1982; Freeman 1989; Hammer 1979; Hummon & Carly 1993; Wellman 1988) investigates sets of relationships that exist in the complex interactions of social members in the context of the social system in which they act (Erickson 1988; Lorrain & White 1971; Scott 2000). Analogous to emergent properties and phase changes common in complex systems, such social transitions result from network topology and information exchange between connected network members (Bura *et al.* 1995; Fliedner 2001; Holyst *et al.* 2000; Klüver & Schmidt 1999; Schecter 2002). Network models provide a natural and effective means of representing hierarchical levels in social systems (Fliedner 2001; Hammer 1979; Hummon & Carly 1993; Sanil *et al.* 1995; Wassermann & Faust 1995).
Graph theory (a mature research discipline) can be used to draw maps or topologies of social structures including random graphs, scale-free *al.* 2000; Albert & Barabási 2001; Barrat & Weigt 2000; Doreian 1979; Erdös & Rènyi 1960; Jeong *et al.* 2000; Wassermann 1980; Watts & Strogatz 1998). The dynamic behavior of multi-agent network systems is typically modeled using general rules that describe the behavior of the agents, and topological rules that describe the patterns by which agents are interconnected and communicate (Klüver & Schmidt 1999).

Multi-agent simulations allow us to develop theory, demonstrate robust characteristics and observe the mechanisms behind unexpected, novel, emergent behavior (Brassel *et al.* 1997; Doreian & Stokman 1997; Freeman 1989; Troitzsch 1997; Wellman & Berkowitz 1988. We can investigate patterns that emerge from the interaction of explicitly defined states of individual agents and the causal processes that change these states over time (Deadman & Gimblett 1994; Fararo & Hummon 1994; Hanneman 1995; Itami *et al.* 2000) providing the capacity to study the complexity of these systems *in silico,* when real-world investigation is impractical, improbable, or impossible (Conte Hegselman & Terna 1997; Gilbert & Triotzsch 1999; Troitzsch 1998).

Clustering

Clustering appears as a phenomenon in diverse systems. It appears to be a common mechanism for coping with complexity. The formation of hierarchies of clusters reduces internal interactions and constrains behavior (Green 2002). Much has yet to be explained about how clusters emerge in multi-agent systems. Social groups can be described as clusters, alliances and networks, where common ideals, interests, and the like link individuals together (Lee 1980). They are formed from, and are maintained by, the patterns of connectivity and information exchange between members (Gilbert 1997). These patterns will influence the collective opinion of the network.

Previous studies by Stocker *et al*. (2001, 2002, 2003) focus on a binary choice of state, that is, either agreement (yes) or disagreement (no) about an issue. A significant research question concerns how a range of different opinions or ideas among a group of individuals connected by different network structures will affect collective or global opinion as members interact over time.

Individual states, social structure, and global opinion

In real-world situations, public opinion is usually diverse and spread across many different ideas, attitudes, and preferences. From time to time, there occurs a coalescing of public opinion towards main ideas that strongly resist changes over time (Schecter 2002), even though each individual makes a choice from several different ideas (Lomborg 1997). A relatively new area of research, Memetics, suggests that certain characteristics of ideas themselves influence selection (Aunger 2002; Blackmore 1999; Brodie 1996; Gladwell 1999; Lynch 1996; Marsden 2000).

Social network simulation research supports that group opinion is influenced by direct contact and communication between peers (Stocker *et al.* 2001, 2002, 2003). Ideas change depending on a susceptibility to attack from other ideas and the structure of connections between individuals (Hales 1998). Individual influence in the course of social transition is an important determinant of public opinion (Burt 1987). Public opinion change is dependent on the exchange of information between connected individuals (Nowak & Lewenstein 1996).

Social comparison stabilizes agreement of opinion (Granovetter 1978; Werner & Davis 1997) and depends on the nature of individuals and their relationships (Erickson 1988). However, diversity means that disagreement may also result in the formation of sub-groups whose members share similar points of view (Doreian & Stokman 1997). These influences have a critical effect on the dynamics of a social system (Nowak & Lewenstein 1996). Smaller disenfranchised or "fringe" groups can collect around radical opinions or ideas that are not representative of the majority.

In this study, A multi-agent simulation of specific network structures is used to represent the patterns of connection and communication between interacting agents. In what is essentially a network diffusion simulation (Valente 2005; see also Becker 1970; Rogers 1958, 1995) this simulation uses multiple instead of binary choices. The following questions are addressed: 1) how do individual states and membership of different network structures influence global opinion in a social network? and 2) do individual members tend to form a cluster around particular states or ideas so that sub-groups emerge?

Methods

In the simulation, 100 nodes are connected in three different network structures to represent patterns of connectivity and communication. Networks are static (that is, links and population remain the same) and network parameters are shown in Table 1. Each node is randomly initialized with a choice of state from 2 to 10 representing different issues or ideas. Each node is also randomly initialized with values for levels of influence and susceptibility (between 0.0 and 1.0). The result of interactions between nodes over time is observed to determine aggregation of nodes around particular states.

Table 1. Ranges of Values Assigned to Parameters in the Simulation for Each Network Type

	Hierarchy	Random	Scale-free
Network Parameters	2 to 10 (layers)	0.01 to 0.5 (connectivity)	0.0 to 1.0 (constant) 1.6 to 4.0 (exponent)

Nodes in the simulation change states asynchronously as do individuals in real world networks (Harvey & Bossomaier 1997; Cornforth *et al*. 2001, 2002). There are 10 runs of 1000 time steps for each combination of the network structure parameters.

The *expected number of nodes* (E) that adopt a given state at initialization will approximate to the *population* (N) divided by the *number of states* (S). For example, for 100 nodes with a choice of 5 states, we expect 20 nodes to adopt each state. Comparison of this expected value with the *average maximum number of nodes that adopt each state* after T iterations (time steps) provides a reliable measure of the relative degree to which nodes have adopted particular states as a result of their interaction. This comparison, expressed as an *Adoption Ratio* (AR) is defined by:

$$AR = \frac{\frac{1}{T}\left[\sum_{t=1}^{T}\max\left(n_1(t), n_2(t), \ldots n_i(t)\right)\right]}{\frac{N}{S}}$$

[Eq 1.]

Where $n_i(t)$ is the number of nodes adopting each state i at time t, N is the number of nodes in the population, T is the total number of time steps and S is the number of states available.

Results

Selected results below describe the behavior of hierarchy, random and scale-free networks and demonstrate that cohesion (the number of nodes in the same state) varies for the number of different states available to each node over time. As the number of states available increases, the maximum number of nodes adopting each state decreases, regardless of network structure.

There is a significant change of node "loyalty" to specific states where the maximum number of nodes adopting one or another state is not consistent. That is, aggregation of the nodes around specific states occurs to varying degrees. However, in each experiment only two or three states emerge as having the largest number of nodes adopting that state. This suggests that clustering is an emergent feature or principle of social structures independent of the parameters associated with network structure.

Of interest in hierarchy networks, is the critical change around a depth of 5 to 6 hierarchy layers (Figure 1). In the random network, there is evidence of critical behavior at a connectivity level of 0.25 to 0.3 (see Dunbar 1995; Wellman 1988) (Figure 1). Clustering of nodes around particular states is more evenly spread across the different states and node "loyalty" is less evident than in hierarchy networks.

As the scale-free constant (Z), the scale-free exponent (λ) and the number of states vary, the clustering of nodes around one to three states is less pronounced than for either the hierarchy or random structures. At a scale-free constant of 0.25, as the scale-free exponent and the number of states are varied, the maximum number of nodes adopting a state is only affected after the scale-free exponent reaches a value of 2.8. Interestingly, when the number of states is around 4, a significant reduction in the largest cluster size occurs for values of the scale-free exponent above 2.8. As the scale-free exponent moves above 2.8, a linear increase (approximate) in the maximum number of nodes adopting particular states occurs.

When the scale-free exponent is 2.2 and the scale-free constant and the number of available states are varied, the largest cluster is affected when the scale-free constant reaches 0.5. Again, when the number of states is around 4, a significant reduction in the largest cluster size occurs for values of the scale-free constant above 0.5 (Figure 1).

These preliminary results are somewhat expected. It is logical that when the number of choices available to a group is increased, the maximum number of individual's choosing a particular state will necessarily decrease. Likewise, it is apparent that, at key values for each network's structural parameters, there is change in the global behavior of the network - also an intuitive result.

Figure1. Maximum Cohesion in a Population of 100 Nodes for Hierarchy, Random, & Scale-free Networks

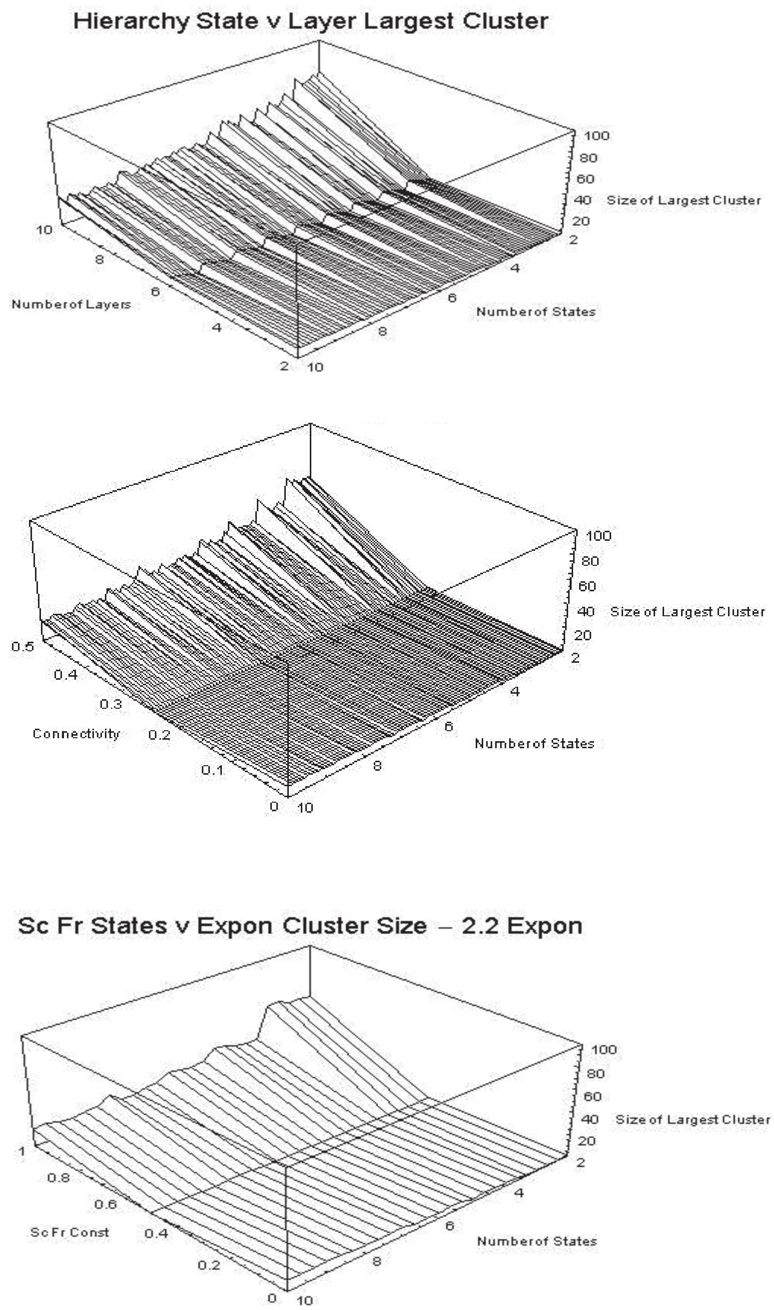


Figure 1. The surface plots show the relationship between the number of states that nodes can adopt and the key parameters of each network structure (viz, hierarchy layers, random connectivity and scale-free exponent/constant).

However when we examine the results from calculating the Adoption Ratio *AR* (Equation 1) over time, it is evident that counter-intuitive behavior is occurring. The behavior manifests as an increase in the *AR* as the number of states available increases. This occurs consistently for each of the hierarchy, random, and scale-free network structures.

In the hierarchy network, *AR* is consistently greater than 1.1 and varies between 1.1 and 2.2, showing clustering behavior away from the initialized state of the model. As the number of states is held constant and the depth of the hierarchy is varied, and the number of layers increases, *AR* remains fairly constant with a peak at 5 to 6 hierarchy layers. Counter-intuitively, as the number of layers is held constant and the number of states available is varied, the Adoption Ratio increases (Figure 2 – red squares).

For the random network, as the number of states is held constant and connectivity is varied *AR* remains within the range 1.1 to 1.8, with increased activity around critical connectivity of 0.25 to 0.30. When connectivity is held constant and the number of states is varied, *AR* increases with the number of states available with peaks at 5 and 7 states, a counter-intuitive result, indicating cluster size is dependent on number of states (Figure 2 – yellow triangles).

With the scale-free network structure, *AR* shows a steady increase as the number of available states increases with a peak at 6 states (Figure 2 – blue diamonds), indicating that cluster size increases with the number of available states. The similarity to hierarchy and random network behavior is evident.

Figure 2. Adoption Ratios by the Number of States in Hierarchy, Random, and Scale-free Network Structures

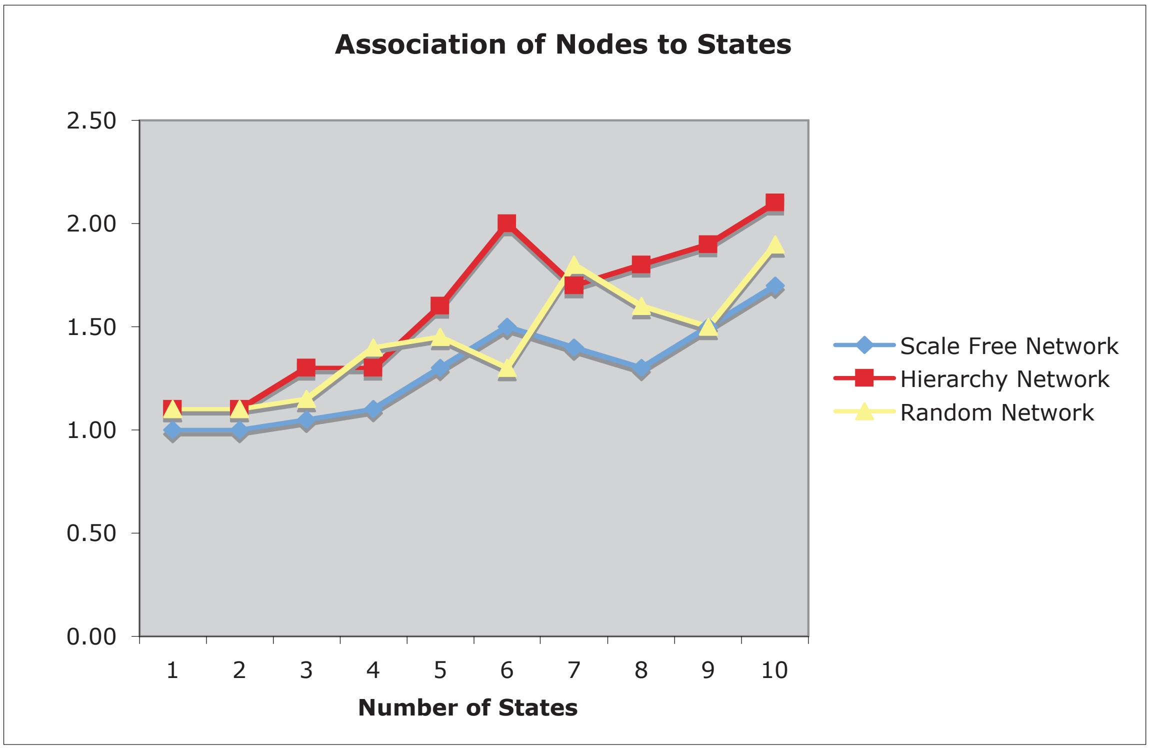


Figure 2. *Graph of the Adoption Ratio (AR) where the Y Axis shows AR and the X Axis shows the number of states in each of the hierarchy, random, and scale-free network structures. It shows the increase of AR with increasing number of states available.*

Discussion and Conclusions

Society comprises individuals who are connected by their involvement in work, social organisations, sporting clubs, religious communities, and so on. As members of more than one group they are influenced by the opinions, attitudes, and ideas of other members of the groups to which they belong. The boundaries of these groups often enclose a small to medium population of around 100 to 150 (Dunbar 1992, 1993; Wellman 1988). The manner in which members are connected will also vary.

This simulation suggests that network structure has an impact on the formation of public opinion in groups of social members that share common ideals, attitudes, or opinions. There is criticality with respect to parameters associated with network structure. In different social structures: (1) a majority of the population will change state from a large range of ideas to form aggregates, groups or clusters around on two to three preferred ideas, and, (2) clustering is dependent on the parameters associated with the patterns of connectivity between peer nodes (the structure of the networks). Clusters emerge as a result of a choice of states, although the maximum number of nodes that adopt each specific state reduces as the number of available states increases. This confirms an intuitive understanding that the more choices there are, the more difficult it is to make a choice. However, the *Adoption Ratio* (*AR)* demonstrated counter-intuitive behaviors. Regardless of the type of network, as the number of states available increases so did the Adoption Ratio, which demonstrates universality. One possible explanation (not yet confirmed) for this phenomenon is that some form of positive or reinforcing feedback is occurring amongst individuals in the social structure. As the number of nodes that adopt a specific state increases from the initial state above that expected, those nodes exert a converting force on the other nodes to which they are connected, regardless of network structure. However, there is a tension between this effect and the individual nodes' levels of influence and susceptibility. This tension provides the main constraint against the connected nodes changing state, thus explaining why the whole population is not converted. There is also criticality in respect of the parameters inherent in the different network structures. These factors suggest that there are general principles that apply to the formation of sub-groups within fixed populations, regardless of the structure to which they belong.

There are implications from this research for the management of change and survival of groups whose members will be connected to sources of different ideas, attitudes and opinions. Future research will focus on the behavior of dynamic networks where the network structure is influenced by the addition and deletion of nodes and links.

References

Albert, R. and Barabási. A.L. (2001). Statistical mechanics of complex networks. *arXiv:cond-mat/0106096*, 1, 1-54.

Albert, R., Jeong, H., and Barabási. A.L. (2000). Error and attack tolerance of complex networks. *Nature*, 406, 378-382.

Aunger, R. (2002). *The electric meme: A new theory of how we think.* New York: The Free press (Simon and Schuster Inc.).

Bak, P. and Sneppen, K. (1993). Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters*, 71 (24), 4083-4086.

Barabási. A.L. and Bonabeau, E. (2003) Scale-free networks. *Scientific American*, May, 50-59.

Barak, A. and Weigt, M. (2000). On the properties of small-world network models. *Eur. Phys. J. B*., 13, 547-560.

Becker, M.H. (1970). Sociometric location innovativeness: Reformulation and extension of the diffusion model. *American Sociological Review*. 35 (2): 267-282.

Berkowitz, S. D. (1982). *An introduction to structural analysis* Toronto: Butterworths.

Blackmore, S. (1999). *The meme machine.* New York: Oxford University Press Inc.

Brassel, K-H Möhring, M., Schumacher, E. and Troitzsch K.G. (1997) Can agents cover the world?. in R Conte *et al.* (eds) *Simulating social phenomena*. Heidelberg: Springer-Verlag. 57-72.

Brodie, R. (1996). *Virus of the mind: The new science of the meme.* Seattle: Integral Press.

Bura, S., Guérin_Pace, F., Mathian, H. Pumain, D. and Sanders, L. (1995). Cities can be agents too: A model for the evolution of settlement systems. in N Gilbert and R Conte (Eds) *Artificial societies: The computer simulation of social life.* London: University College London Press. 86-102.

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

Conte, E.R., Hegselmann and Terna P (1997) Introduction: Social simulation - A new disciplinary synthesis. in R Conte *et al.* (Eds) *Simulating social phenomena*. Heidelberg: Springer-Verlag, 1-10.

Cornforth, D. Green, D.G. Newth, D. and Kirley, M. (2001). Ordered asynchronous processes in natural and artificial systems. in Peter Whigham *et al.* (Eds) *Proceedings of the Fifth Japan-Australia Joint Workshop on Intelligent and Evolutionary Systems.* Dept of Information Science, University of Otago Dunedin NZ. 105-112.

Cornforth, D. Green, D.G. Newth, D. and Kirley, M. (2002). Do artificial ants march in step? Ordered asynchronous processes and modularity in biological systems. in R K Standish *et al.* (Eds) *Proceedings of the Eighth International Conference on Artificial Life*}. Sydney. 28-32.

Deadman, P. and Gimblett, H. R. (1994). The role of goal-oriented agents in modeling people-environment interactions in forest recreation. *Mathematical and Computer Modelling.* 20 (8).

Doreian, P. (1979). On the evolution of group and network structure. *Journal of Social Networks.* 2: 235-252.

Doreian, P. and Stokman F.N. (1997). *The dynamics and evolution of social networks.* New York: Gordon and Breach.

Dunbar, R.I.M. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution.* 20: 469-493.

Dunbar, R.I.M. (1993). Co-evolution of neo-cortex size, group size and language in humans. *Behavioral and Brain Sciences.* 16 (4): 681-735.

Dunbar, R.I.M. (1995). Neocortex size and group size in primates: A test of the hypothesis. *Journal of Human Evolution.* 28: 287-296.

Dunbar, R.I.M. (1996). *Grooming, Gossip and the Evolution of Language.* London: Faber.

Erdos, P. and Rényi, A. (1960). On the evolution of random graphs. *Mat. Kutato Int. Kozl.* 5: 17-61.

Erickson, B.H. (1988). The realtional basis of attitudes. in Barry Wellman and S D Berkowitz (Eds) *Social Structures: A Network Approach.* New York: Cambridge University Press. 99-121.

Fararo, T.J. and Hummon, N.P. (1994). Discrete event simulation and theoretical models in sociology. *Advances in Group Processes.* 11: 25-66.

Fleidner, D. (2001). Six levels of complexity: A typology of processes and systems. *Journal of Artificial Societies and Social Simulations*}. 4(1): 4.

http://www.soc.surrey.ac.uk/JASSS/4/1/4.html.

Freeman, L. (1989). Social networks and the structure experiment. in L C Freeman *et al.* (Eds) *Research methods in Social Network Analysis.* Fairfax VA: George Mason University Press. 11-40.

Gilbert, N. (1997). Emergence in social simulation. in Nigel Gilbert and Rosaria Conte (Eds) *Artificial Societies: The computer simulation of social life.* London: University College London (UCL) Press. 144-156.

Gilbert, N. and Troitzsch, K.G. (1999). *Simulation for the Social Scientist.* Buckingham, UK: Open University Press.

Gladwell, M. (1999). *The Tipping Point.* London: Little Brown.

Goldspink (2002). Methodological implications of complex systems approaches to sociality simulation as a foundation for knowledge. *Journal of Artificial Societies and Social Simulations.* 5 (1): 3.

http://www.soc.surrey.ac.uk/JASSS/5/1/3.html.

Granovetter, M. (1978). Threshold models of collective behavior. *American Journal of Sociology.* 83 (6): 1420-1443.

Green, D.G. (1993). Emergent behavior in biological systems. in D G Green & T R J Bossomaier (Eds) *Complex Systems 92: - From Biology to Computation.* Amsterdam: IOS Press. 24-35.

Green, D.G. (2002). Hierarchy, complexity and agent models. in *Our Fragile World.* Paris: UNESCO.

Hales, D. (1998). An open mind is not and empty mind: Experiments in the meta-noosphere. *Artificial Societies and Social Simulations.* 1(4), 1-26.

Hammer, M. (1979). Predictability of social connections over time. *Journal of Social Networks.* 2: 165-180.

Hanneman, R.A. (1995). Simulation modelling and theoretical analysis in sociology. *Social Perspectives.* 38 (4): 457-462.

Harvey, I. and Bossomaier, T.R.J. (1997). Time out of joint: Attractors in asynchronous boolean networks. in P Husbands & I Harvey (Eds) *Proc. Fourth European Conference on Artificial Life.* 67-75.

Holland, J. (1995). *Hidden order: How adaptation builds complexity.* Reading MA: Perseus Books.

Holyst J.A, Kacperski, K. and Schweitzer, F. (2000) Phase transitions in social impact models of opinion formation. Elsevier Preprint arXiv:cond-mat/ooo4026}. 1 (1): 1-21.

Horgon, J. (1995). From complexity to perplexity. *Scientific American*}. June. 74.

Hummon, N.P. and Carley, K. (1993). Social networks as normal science. *Social Networks.* 15: 71-106.

Itami, R.M. MacLaren, G.S., Hirst, R.J., Raulings, R.J. and Gimblett, H.R. (2000). RBSIM2: Simulating human behavior in National Parks in Australia: Integrating GIS and intelligent agents to predict recreational conflicts in high use natural environments. *4th International Conference on Integrating GIS and Environmental Modelling (GIS/EM4): Problem and research needs.* Banff, Alberta Canada.

Jeong, H., Tombor, B., Albert, R., Oltavi, Z.N. and Barabási, A.L. (2000). The large scale organisation of metabolic networks. *Nature*}. 407: 651-654.

Kluver, J. and Schmidt, J. (1999). Topology, metric and dynamics of social systems. *Journal of Artificial Societies and Social Simulations.* 2 (3): 7.

http://www.soc.surrey.ac.uk/JASSS/2/3/7.html.

Langton, C.G. (1990). Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D.* 42: 12-37.

Lee, T.R. (1980). The resilience of social networks to change in mobility and propinquity. *Social Networks*}. 2: 423-435.

Lomborg, B. (1997). Simulating multiparty systems. In R. Conte *et al. Simulating Social Phenomena.* Heidelberg: Springer-Verlag. 335-341.

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

Lynch, A. (1996). *Thought contagion: How belief spreads through society.* New York: Basic Books (Perseus Books Group).

Marsden, P. (2000). Mental epidemics. *New Scientist.* 2237: 46-47.

Nowak, A. and Lewenstein, M. (1996). Modelling social change with cellular automata. in R Hegselmann (Ed) *Modelling and simulation in the social sciences from the philosophy of science point of view.* Netherlands: Kluwer Academic Publishers. 249-285.

Rogers, E.M. and Beal, G.M. (1958). The importance of personal influence in the adoption of technological changes. *Social Forces*. 36 (4): 329-335.

Rogers, E.M. (1995). *Diffusion of innovations.* New York: Free Press.

Sanil, A. Banks, D. and Carley, K. (1995). Models for evolving fixed node networks: Model fitting and model testing. *Social Networks.* 17: 65-81.

Schecter, B. (2002). The new physics of peer pressure. *New Scientist.* August: 42-45.

Scott, J. (2000). *Social Network Analysis: A Handbook.* London: Sage Publications.

Stocker, R., Green, D.G. and Newth, D. (2001). Consensus and cohesion in simulated social networks. *Journal of Artificial Societies and Social Simulations.* 4 (4): 5.

http://www.soc.surrey.ac.uk/JASSS/4/4/5.html.

Stocker, R., Cornforth, D. and Bossomaier, T.R.J. (2002). Network structures and agreement in social network simulations. *Journal of Artificial Societies and Social Simulations.* 5 (4): 3.

http://www.soc.surrey.ac.uk/JASSS/5/4/3.html.

Stocker, R., Cornforthm D. and Green, D.G. (2003). A simulation of the impact of media on social cohesion. *Journal of Advances in Complex Systems.* 6 (3): 349-359.

Troitzsch, K.G. (1997). Social science simulation - Origins, prospects, purposes. in R Conte *et al.* (Eds) *Simulating Social Phenomena.* Heidelberg: Springer-Verlag. 40-54.

Valente, T. W., Unger, J.B. and Johnson, C.A. (2005). Do popular students smoke? The association between popularity and smoking among middle school students. *Journal of Adolescent Health* 37 (4): 323-329.

Wasserman, S. (1980). Analysing social networks as stochastic processes. *Journal of the American Statistical Association.* 73 (370): 280-294.

Wasserman, S. and Faust, K. (1995). *Social network analysis: Methods and applications.* Cambridge: Cambridge University Press.

Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature.* 393: 440.

Watts, D.J., Dodds, P.S., and Newman, M.E.J. (2002). Identity and search in social networks. *Science.* 296: 1302-1305.

Wellman, B. (1988). Structural analysis: From method and metaphor to theory and substance. in B Wellman and S D Berkowitz (Eds) *Social Structures : A Network Approach.* Cambridge: Cambridge University Press. 19-61.

Wellman, B. and Berkowitz, S.D. (1988). Introduction: Studying social structures. in B Wellman and S D Berkowitz (Eds) *Social Structures : A Network Approach.* Cambridge: Cambridge University Press. 1-18.

Werner, G.M. and Davis, J.N. (1997). Cooperation without memory. in R Conte *et al.* (Eds) *Simulating Social Phenomena.* Heidelberg: Springer-Verlag. 179-187.

Co-Citation of Prominent Social Network Articles in Sociology Journals: The Evolving Canon

David Lazer[1]
Ines Mergel[2], **Allan Friedman**[1]

[1]*Kennedy Kennedy School*
Harvard University

[2]*Maxwell School of Citizenship and Public Administration*
Syracuse University

Abstract

Social network analysis has been a particularly hot area across the social (and some non-social) sciences. How has this growth, in turn, affected the field of social network analysis within sociology, the discipline which has served as the primary home of social network analysis over the last several decades? In order to answer this question, we examined the citation patterns of the social network papers in the two leading general sociology journals, the *American Sociological Review* and the *American Journal of Sociology*, from 1990-2005, focusing on the body of literature that was cited by at least two social network papers in a given year. We produced two network snapshots of the social network canon during this period.

These analyses reveal a combination of great change and substantial continuity. There was a substantial increase in interest in social networks in sociology throughout this period, and, in particular, an enormous rise in interest in small world issues, coupled with the abrupt entry of mathematicians and physicists into the sociology social network canon. However, during this entire period Granovetter's work remained squarely at the center of the canon, with Granovetter (1973) as the most cited piece at both the earlier and later snapshots.

Please address correspondence to David Lazer at david_lazer@harvard.edu

Introduction

The study of networks has been one of the major growth areas within scholarly research over the last decade. Particularly striking has been the growth of research on networks within physics, dating from Watts and Strogatz's (1998) work on small world networks. What impact has this growth had on the study of networks within sociology, a discipline that has served as the primary home for the study of networks for the last several decades? In this paper, we examine the co-citation pattern of papers published in the two leading generalist journals within sociology, the *American Sociological Review* (ASR) and the *American Journal of Sociology* (AJS), for 1990-2, 2000, and 2005. A co-citation is a shared reference of two articles to a third source. The list of co-cited references of social network research within ASR and AJS for a given year offers a rough measure of what the field collectively believes is within the canon: those sources worthy of attention and acknowledgement. How has the content of the canon evolved over the years? In particular, what impact has the work within physics had on the study of social networks within sociology? What is the underlying structure of the canon? For example, is there a common core of sources that all social network articles cite? Or does the field have a more decentralized structure?

Our analysis reveals a combination of great change and substantial continuity in the field. There was a major increase in interest in social networks in sociology during this period, and, in particular, an enormous rise in interest in small world issues, coupled with the abrupt entry of mathematicians and physicists into the sociology canon. However, across all periods, Granovetter's work remained squarely at the center of the canon, with Granovetter (1973) as the most cited piece in both the early and late periods.

Co-citation analysis

The list of citations within a published article offers a glimpse into what is considered the *canon* at a particular point in time of the field, reflecting the collective wisdom of the author, editor, and referees as to what prior research acts as the foundation for the findings of that article. The list of citations for a particular article will certainly reflect the idiosyncrasies of that particular author, and details of the article's subject area. However, the *body* of articles published in a given year reflects a communal consensus as to what the collective research agenda is, and, in particular, what prior research is worth paying attention to. We therefore used the concept of *co-citation analysis* from bibliometrics (White/Griffith 1981). A co-citation occurs when two articles share a reference. Co-citation analysis is used in different ways: it can help to identify so-called "invisible colleges" in forms of clusters of authors who cite similar references; it can also detect emerging trends within a research field or shows bridges among research disciplines (see Table 1 for examples).

Table 1. Examples of Co-citation Analysis

Study	Field	Authors
Co-Authorship in Management and Organization Studies	Organizational Behavior	(Acedo/Barroso et al. 2006)
Search for invisible colleges	Methodological evaluation	(Gmuer 2003)
Bridges between research disciplines	Information Science	(Karki 1996)
Intellectual Development of MIS: clusters/invisible colleges	Management Information Science	(Culnan 1987)

We contrast co-citation research with the recent work on co-authorship networks that illuminates the emerging structure of collaboration among academics within different disciplines (Newman 2001; Barabasi/Jeong et al. 2002; Newman 2004). While co-authorship networks reflect the structure of collaboration within a research community, co-citation networks reflect the *structure of attention* within a research community—that is, what prior research is worth paying attention to (de Solla Price 1965).

The standard procedure to conduct a co-citation analysis is presented by McCain (1990), see also description of the process in Ahlgren/Jarneving et al. (2003, p. 550). In our study, we have adopted the following procedures:

- Identify articles that are primarily focused on the study of social networks in the top two generalist sociology journals—AJS and the ASR—for 1990-92, and in 2000 and 2005. (We aggregate 1990-92 in order to produce a list of seed articles of a comparable size as 2005, allowing useful structural comparisons.) We therefore do not claim to offer an overall picture of social network analysis as a field but rather of social network analysis within sociology. Articles were selected by hand-coding the abstracts (see Appendix A).

- The citations from each article were collected, where for each period studied, we eliminated cited literature that was only referenced by one article. In short, inclusion in the canon requires a minimum of "two votes" from top journals.

The resulting graphs offer a picture of the evolution of structure of the sociologically-based social network analysis as a field.

Social Network articles in ASR and AJS

A comparison across the years highlights the increase in the quantity of research on social networks. For example, while there were 20 articles published 1990-92 that examined social networks, there were none in 1995, 7 in 2000, and 14 publications in 2005. This increase is consistent with, but not as dramatic as, findings described by Borgatti and Foster (2003) in their review of the applications of social network analysis in the field of organizational behavior. Perhaps this contrast reflects that social network analysis was starting from a higher base in sociology. In addition, across all of the years, AJS appears to be far more likely to provide a venue for social network research than ASR. Figure 1 portrays the co-citation structure for ASR and AJS from 2005. The seed articles are circles, and the cited papers and books are squares. We also distinguish between articles from ASR (yellow) and AJS (red). We note that AJS featured a "Special Issue on Computation" in 2005 that prominently featured social network research, contributing five of the fourteen articles from that year. We denote these articles with a dark outline and consider them further below. The black squares are from the social sciences, and grey squares from mathematics and physics.

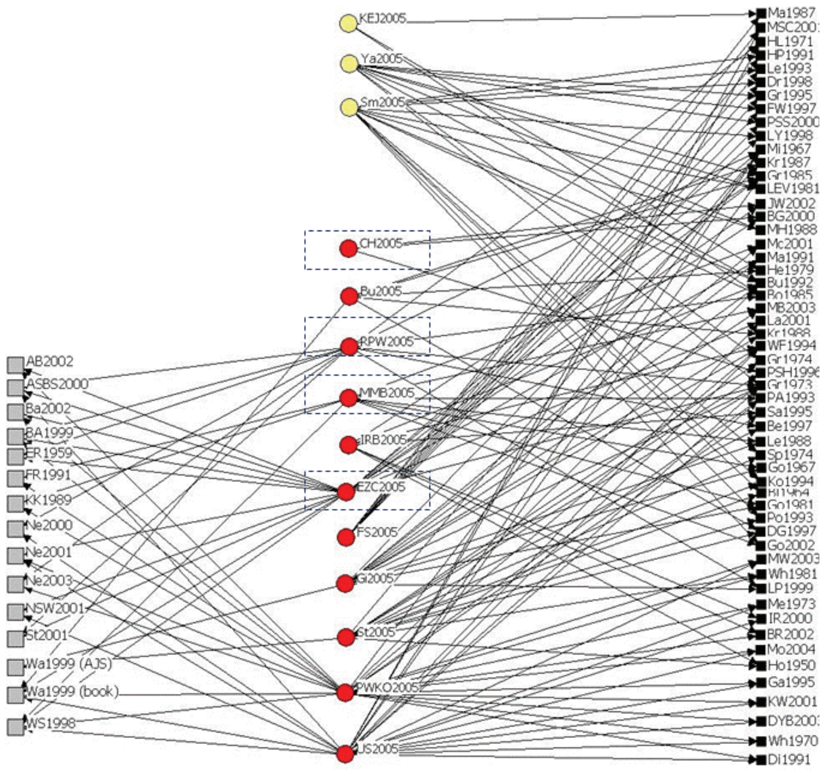


Figure 1. Co-citation Patterns 2005[1]

A comparison over the years also reflects the fact that the average number of total citations in each article grew during this period. Figure 2 highlights the growth of the number of citations where the average social network article had 47% more references in 2005 than in the early 1990s. (This partly reflects the secular trend during this period toward the inclusion of more references where the average non-network articles in AJS/ASR experienced approximately a 26% increase in number of references.)

---

1 An earlier version of this figure appeared in Heyman (2006: p. 606).

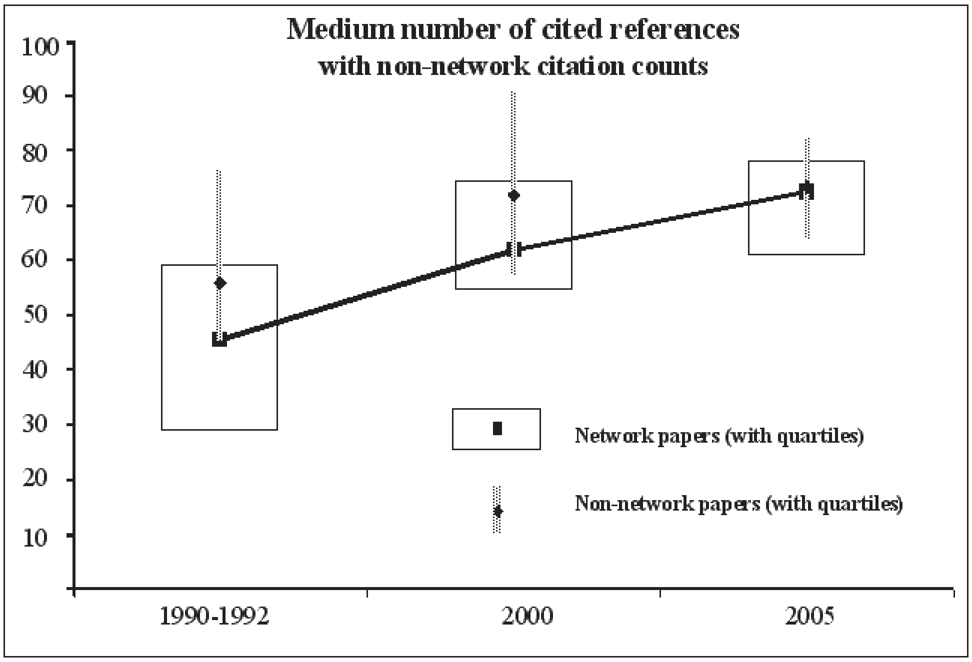


Figure 2. Comparison of Total Number of Articles Cited by Year (1990-2005) (overlaid with non-network citation pattern distribution)

The data embedded in these figures also reveals the "erosion" of older pieces of the canon, i.e., how the probability of being cited declines over time. Figure 3 plots the proportion of papers cited for each of the three periods against the year of the citation. It generally appears that attention is maximized about four years after an item is published, dropping about 50% every 5-7 years after that.[2] We would note, however, of the top four articles cited in 2005, one is from the early 1990s, one from the 1980s, one from the 1970s, and one from the 1960s (see Table 2). Figure 3 thus reflects the fact that most articles largely disappear after a decade, but a handful of classics continue to receive citations for long after.

2 We would note that the "decline" for the 2005 data is exaggerated by the fact that there likely more references in the preceding five-year period than any other period that references were observed.

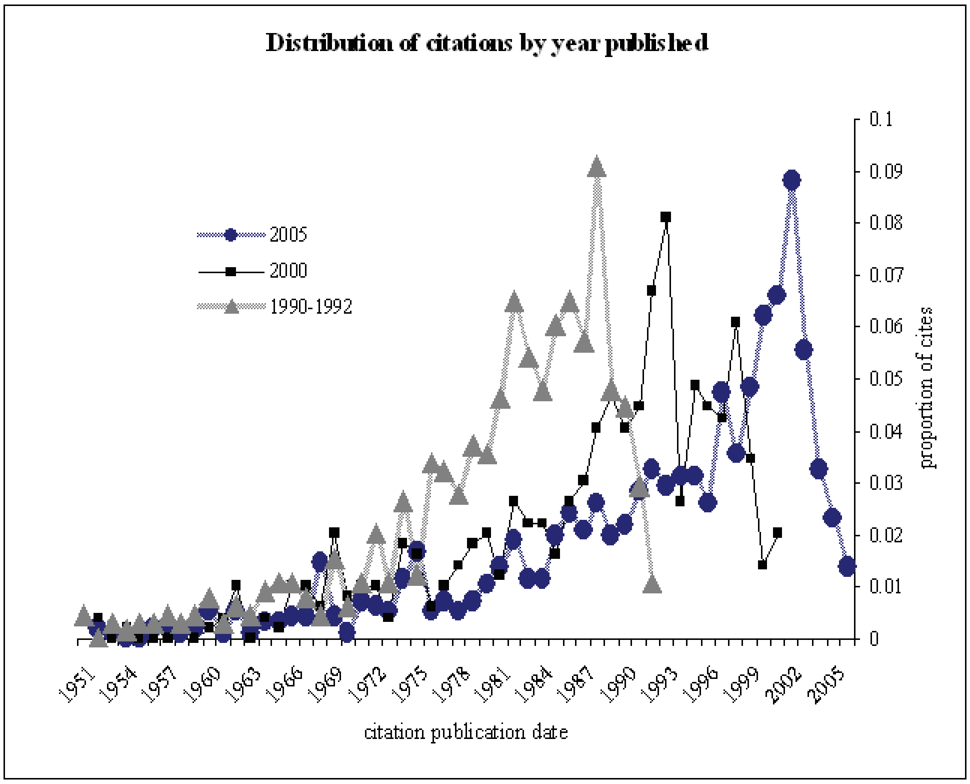


Figure 3. Distribution of Citations by Year Published

Of the 46 articles cited multiple times in 1990-1992, three are cited at least twice in 2000. Notably, Granovetter (1973), Granovetter (1974) and Lin/Ensel et al. (1981) also reappear in 2005, along with four others found in the early 1990s. Of the 20 articles in the canon of 2000, five remain in the 2005 canon. In addition, there were 16 articles from the 1980s or earlier that appeared in the canon in 2005 that did not appear in 2000 or the early 1990s (in comparison, 45 co-cited articles in the early 1990s were from the 1980s or earlier). This certainly reflects the vagaries of the particular articles that happened to appear in the time periods we looked at, where, for example, Krackhardt (1987; 1988) was certainly in the canon during the 1990s. However, it is also clear that some research veins that had faded over the years have returned to prominence - in particular, with respect to “small world” research. For example, Milgram (1967) was not co-cited in the earlier periods, but became the second most cited article in 2005, likely due to its complementarity with Watts (1999) and Watts and Strogatz (1998). Similarly, fairly old work on random graphs was also resurrected - e.g., Erdös and Renyi (1959).

Other items, at the center of the canon in the early 1990s - e.g., Homans (1974) and Fischer (1982a; 1982b) - essentially disappeared. These shifts, in part, reflect the movement of the field into different areas. In addition, concepts such as network centrality that used to demand a specific citation are now accepted generic metrics. And, of course, the nature of academic trends demand a fresh set of underpinnings for each wave of analysis in the forward progress of the field. Appendix B presents the articles and books that were cited at least twice in any of the time periods we examined.

The most striking change in the canon, however, is what Bonacich (2004) termed "*the invasion of the physicists*." We identify the recent work by mathematicians and physicists, which made a rapid entry into the canon after 2000—from *no* co-citations in the early 1990s and 2000 to 37 (23% of the total) of all co-citations in 2005. The AJS special issue on computation drives a disproportionate number of these citations. A full 28 of the 37 (~75%) of those co-citations are from the special issue, where three of the four heavy science citers are found in this issue.

Table 2 provides an additional sense of evolution of the field: it includes a list of the most co-cited articles during these three periods; and a list of the most co-cited authors. The mathematicians/physicists are highlighted in italics.

Table 2. Most Cited Articles and Author by Year

Most cited articles by year		
1990-92	**2000**	**2005**
Granovetter 1973 (4) Fischer 1982 (4) Homans 1974 (4)	Fernandez and Weinbert 1997 (3) (Multiple articles at 2)	Granovetter 1973 (6) Granovetter 1985 (5) Milgram 1967 (5) Burt 1992 (5) *Watts and Strogatz 1998 (4)* *Watts 1999 (4)* Wasserman and Faust 1994 (4)
Most cited authors by year		
1990-92	**2000**	**2005**
Granovetter (13) Marsden (11) Fischer (6) Lin (5) Freeman (5)	Granovetter (6) (No one else above 3)	Granovetter (14) *Watts* (*12*) *Newman* (*9*) *Strogatz* (*8*) *Barabasi* (*7*)

The table highlights the striking dominance of Granovetter in this field across all periods. However, equally remarkable is that four of the five most cited authors in 2005 are mathematicians and physicists.[3]

In order to get another view of the evolution of the canon, we converted the 2-mode seed-article-by-co-citation matrix into a 1-mode reference-by-reference affiliation matrix. That is, references A and B are are assumed to be linked if they were on the same list of references (and the more that they appear together, the more strongly they are linked). Figure 4 shows the resulting graph using valued data for 1990-1992, where the link strength is based on the number of times two articles appeared on a reference list together. The figure shows only papers with more than one shared reference. There are a number of clusters of references that tend to be cited together where one cluster is dominated by Granovetter's strength of weak ties work (1973, 1974, 1982), and, to a lesser extent, by Wellman (1982, 1988) and Fischer

3 This is not a completely fair comparison in that some of this work was co-authored among these individuals. However, coauthorship among physicists/mathematicians was not obviously higher than sociologists (with the notable exception of Granovetter, who was single author on all of his co-cited pieces, making his dominance in Table 2 all the more striking).

(1982a, 1982b). There are two other clusters, bridged by Yamagishi et al. (1988), one of which is dominated by Homans (1974) and Blau (1964), and the other by Freeman (1979). The canon in the early 1990s, while all connected, looks easily decomposable into multiple areas, the biggest of which focused on the strength of weak ties, the second on exchange, and the third on centrality.

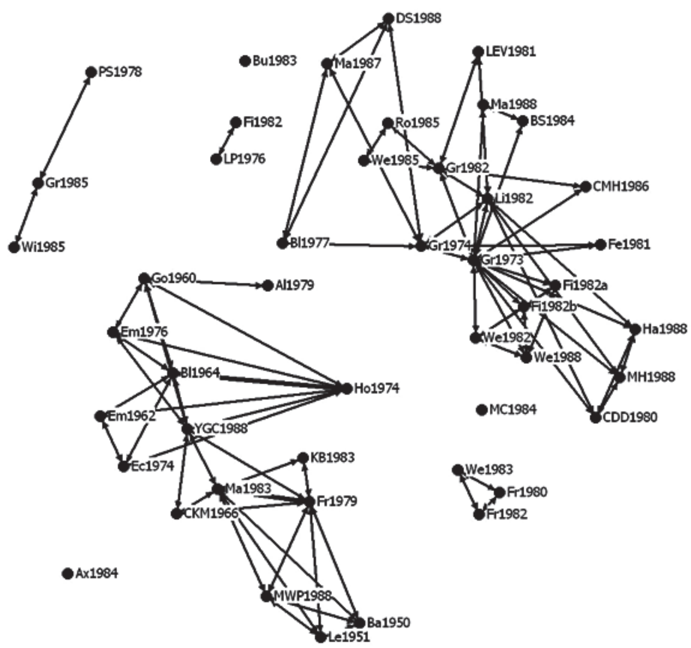


Figure 4. Affiliation Diagram: Subgroups Among Co-cited Articles (1990-1992)

Figure 5 provides the equivalent figures for the 2005 data, with the physics/mathematics articles highlighted as grey squares. The graph of the 2005 data looks dramatically different than the early 1990s. Whereas the 1990s had a number of fairly equally balanced clusters, the 2005 data reveal a clear core-periphery structure, where the core is dominated by a set of strongly connected articles by physicists/mathematicians. This comes through most clearly in Figure 5, where there is one, very large, well-connected component to which all of the physics articles belong, a few small components (e.g., around Goffman), and a number of isolates. In short, there appears to be a core in the reference structure of sociological social network research, to which all of the physicists belong, but also a large diffuse penumbra, which is only loosely connected to the core. This penumbra is connected to the core through a few key articles, such as Granovetter (1973; 1985), Burt (1992) and Milgram (1967).

While we would be hesitant to predict that this core-periphery structure will be reflective of the citation pattern of a more extended period because of the AJS special issue in 2005, it is notable that of the eight social network articles not from the special issue, half cite at least one of the physicists.

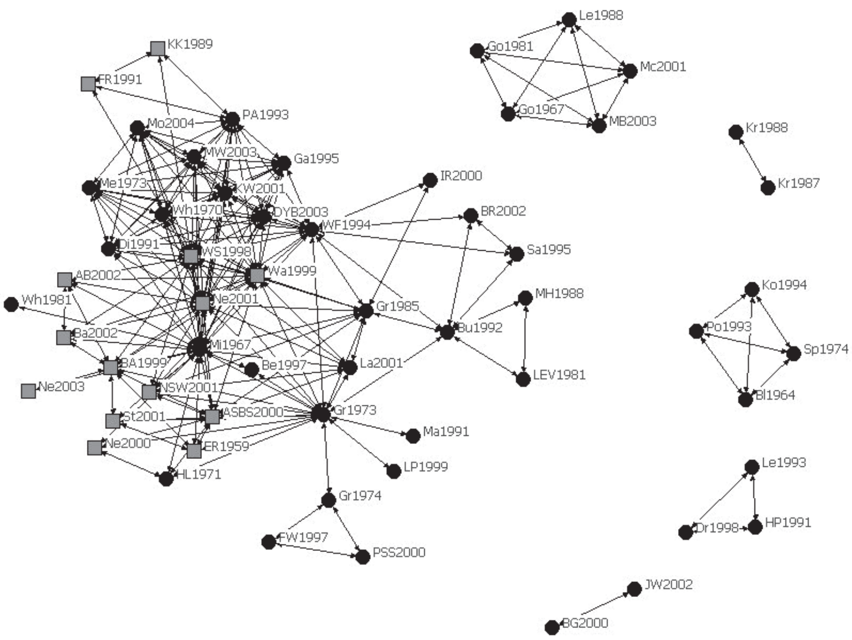


Figure 5. Affiliation Diagram: Subgroups Among Co-cited Articles (2005)

Conclusion

Our co-citation analysis of the social network literature within sociology highlights the rapid evolution of the field in the period 1990-2005. While our analysis suggests that there is a durable core of the field (most notably, around Granovetter's research), it also highlights the rapid entry of the physicists into the canon between 2000 (where no physicists were co-cited) and 2005 (where four of the top five co-cited authors were physicists), and a possible centralization of the field around small-world networks related research.

References

Acedo, F., J., Barroso, C., et al. (2006). Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis. Journal of Management Studies 43: 957-982.

Ahlgren, P., Jarneving, B. et al. (2003). Requirements for a Cocitation Similarity Measure, with Special Reference to Pearson's Correlation Coefficient. Journal of the American Society for Information Science and Technology 54(6): 550-560.

Barabasi, A. L. Jeong, H., et al. (2002). Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications, 311(3-4): 590-614.

Bonacich, P. (2004). The Invasion of the Physicists. Social Networks 26(3): 285-288.

Borgatti, S. P., Everett, M. G. et al. (2002). Ucinet for Windows: Software for Social Network Analysis. Harvard, MA, Analytic Technologies.

Borgatti, S. P., Foster, P. B. (2003). The network paradigm in organizational research: A review and typology. Journal of Management, 29(3): 991–1013.

Burt, R. S. (1992). Structural Holes. Cambridge, Cambridge University Press.

Culnan, M. J. (1987). "Mapping the Intellectual Structure of MIS, 1980-1985: A Co-Citation Analysis." MIS Quarterly 11(3): 341-353.

de Solla Price, D. J. (1965). "Networks of Scientific Papers." Science 149(3683): 510-515.

Erdös, P./A. Renyi (1959). "On Random Graphs. I." Publicationes Mathematicae (Debrecen) 6: 29-97.

Fischer, C. S. (1982a). To Dwell Among Friends: Personal Networks in Town and City. Chicago, University of Chicago Press.

Fischer, C. S. (1982b). What Do We Mean by 'Friend'?, Social Networks 3: 287-306.

Gmuer, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. Scientometrics 57(1): 27-51.

Granovetter, M. S. (1973). The Strength of Weak Ties. American Journal of Sociology, 78: 1360-80.

Granovetter, M. S. (1974). Getting a job: A Study of Contacts and Careers. Cambridge, Mass., Harvard University Press.

Granovetter, M. S. (1985). Economic Action and Social Structure: The Problem of Embeddedness. American Journal of Sociology, 91: 481-510.

Heyman, K. (2006). Sociology: Making Connections. Science, 313(5787): 588-691.

Homans, G. C. (1974). Social Behavior: Its Elementary Forms. New York, Harcourt Brace Jovanovich.

Karki, R. (1996). Searching for Bridges between Disciplines: An Author Co-citation Analysis on the Research into Scholarly Communication. Journal of Information Science, 22(5): 323-334.

Krackhardt, D. (1987). QAP Partialling as a Test of Spuriousness. Social Networks, 9: 171–86.

Krackhardt, D. (1988). Predicting with Networks: Nonparametric Multiple Regression Analysis of Dyadic Data. Social Networks, 10: 359–81.

Lin, N./W. M. Ensel, et al. (1981). Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment. American Sociological Review, 46(4): 393-405.

McCain, K. W. (1990). Mapping Authors in Intellectual Space: A Technical Overview." Journal of the American Society for Information Science and Technology, 41(6): 433-443.

Milgram, S. (1967). The Small World Problem. Psychology Today, 2: 60-67.

Newman, M. E. J. (2001). The structure of scientific collaboration networks. Proceedings of the Natioanl Academy of Sciences of the United States of America, 98(2): 404-409.

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. Proceedings of the Natioanl Academy of Sciences of the United States of America, 101: 5200-5205.

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

Watts, D. (1999). Networks, Dynamics, and the Small-World Phenomenon. American Journal of Sociology, 105(2): 493-527.

Watts, D., Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. Nature, 393: 440-442.

White, H. D., Griffith, C. (1981). Author Cocitation: A Literature Measure of Intellectual Structure. J. ASIS, 32(3): 163-171.F

Appendix A. Seed Articles by Year

1990-1992

Code	ASR	AJS	Seed Reference	Total # of Citations
BO1992			Baum, Joel A. C./Oliver, C. (1992): Institutional Embeddedness and the Dynamics of Organizational Populations, in: *American Sociological Review,* 57/4:540ff.	45
B01990			Bonacich, P. (1990): Communication Dilemmas in Social Networks: An Experimental Study, in: *American Sociological Review,* 55/3 pp. 448-459.	26
Fe1991			Feld, S. L. (1991): Why Your Friends Have More Friends Than You Do, in: *American Journal of Sociology*, 96/6:1464-1478.	8
Fr1991			Friedkin, N. E. (1991): Theoretical Foundations for Centrality Measures, in: *American Journal of Sociology*, 96/6:1478-1505.	61
LGT1992			Lincoln, J. R/Gerlach, M. L./Takahashi, P. (1992): Keiretsu Networks in the Japanese Economy: A Dyad Analysis of Intercorporate, in: *American Sociological Review,* 57/5.	75
Mo1992			Montgomery, J. D. (1992): Job Search and Network Composition: Implications of the Strength-Of-Weak-Tie, in: *American Sociological Review,* 57/5: 586-596.	27
Mo1990a			Moore, G. (1990): Structural Determinants of Men's and Women's Personal Networks, in: *American Sociological Review*, 55/5:726-736.	34
Mo1990c			Molm, L. (1990): Structure, Action, and Outcomes: The Dynamics of Power in Social Exchange, in: *American Sociological Review*, 55/3:427-448.	38
Sh1990			Shrum, W. (1990) Status Incongruence among Boundary Spanners: Structure, Exchange, and Conflict. *American Sociological Review*, Vol. 55, No. 4 pp. 496-511	71
RW1990			Raub, W./Weesie, J. (1990): Reputation and Efficiency in Social Interactions: An Example of Network Effects, in: *American Journal of Sociology*, 96/3:626-655.	23
Ue1990			Uehara, E. (1990): Dual Exchange Theory, Social Networks, and Informal Social Support, in: *American Journal of Sociology*, 96/3:521-558.	65
WW1990			Wellman, B./Wortley, S. (1990): Different Strokes From Different Folks: Community Ties and Social Support, in: *American Journal of Sociology*, 96/3, 558-589.	91
We1991			Wegener, B. (1991): Job Mobility and Social Ties: Social Resources, Prior Job, and Status Attainment, in: *American Sociological Review*, 56/1:60-72.	54

2000

Code	ASR	AJS	Seed Reference	Total # of Citations
FCM2000			Fernandez R. M./Castilla E. J./Moore P. (2000) Social capital at work: Networks and employment at a phone centers, in: *American Journal of Sociology*, 105/5:1288-1356.	61
Ha2000			Hargens, L. L. (2000) Using the literature: Reference networks, reference contexts, and the social structure of scholarship. *American Sociological Review*, 65/6:846-865.	70
HHB2000			Hurlbert, J. S./Haines, V. A./Beggs, J. J. (2000): Core networks and tie activation: What kinds of routine networks allocate resources in nonroutine situations?, in: *American Sociological Review*, 65/4:598-618.	58
PSS2000			Petersen, T./Saporta, I./Seidel, M.-D. L. (2000): Offering a Job: Meritocracy and Social Networks, in: *American Journal of Sociology*, 106/3:763-816.	48
HSS2000			Hedström, P./Sandell, R./Stern, C. (2000): Mesolevel Networks and the Diffusion of Social Movements: The Case of the Swedish Social Democratic Party, in: *American Journal of Sociology*, 106/1:145-173.	51
My2000			Myers, J. D. (2000): The Diffusion of Collective Violence: Infectiousness, Susceptibility, and Mass Media Networks, in: *American Journal of Sociology*, 106/1:173-209.	83
PR2000			Pescosolido, B.A./Rubin, B.A. (2000) The Web of Group Affiliations Revisited: Social Life, Postmodernism, and Sociology, in: *American Sociological Review*, 65/1.	132

2005

Code	ASR	AJS	Seed Reference	Total # of Citations
KEJ2005			Korinek/Entwisle/Jampaklay (2005): Through Thick and Thin: Layers of Social Ties and Urban Settlement among Thai Immigrants, in: *American Sociological Review*, 70:779-800.	72
Ya2005			Yakubovich, V. (2005) Weak ties, information, and influence: How workers find jobs in a local russian labor market. *American Sociological Review*, 70/3:408-421.	
Sm2005			Smith, S. S. (2005): Don't put my name on it": Social Capital Activation and Job-Finding Assistance among the Black Urban, *American Journal of Sociology*; 111:1:1-57.	

Code	ASR	AJS	Seed Reference	Total # of Citations
St2005			Stewart, D. (2005): Social Status in an Open-Source Community, in: *American Sociological Review,* 70:823-842.	49
CH2005			Chang, M-H, Harrington, J.E. (2005) Discovery and Diffusion of Knowledge in an Endogenous Social Network. *American Journal of Sociology*, 110:4.	
Bu2005			Burris, B. (2005): Interlocking Directorates and Political Cohesion among Corporate Elites1, in: The *American Journal of Sociology*, 111/1:249ff.	93
MMB2005			Moody, McFarland, Bender-deMoll (2005: Dynamic Network Visualization, in: *American Journal of Sociology,* 110/4:1206–41.	75
IRB2005			Ingram, P./Robinson, J./Busch, M. L. (2005): The Intergovernmental Network of World Trade: IGO Connectedness, Governance, and Embeddedness, in: *American Journal of Sociology,* 111/3:824ff.	61
EZC2005			Eguiluz, V.M., Zimmerman, M.G., Cela-Conde, C.J., San Miguel, M. (2005) Cooperation and the Emergence of Role Differentiation in the Dynamics of Social Networks, in: *American Journal of Sociology*, 110:4.	74
FS2005			Fernandez, R. M./Sosa, M. L. (2005): Gendering the Job: Networks and Recruitment at a Call Center, in: *American Journal of Sociology*, 111/3:859ff.	84
Gi2005			Gibson, D. (2005): Taking Turns and Talking Ties: Networks and Conversational Interaction, in: *American Journal of Sociology*, 110/6:1561-1597.	64
PKWO2005			Powell, W.W. White, D.R., Koput, K.W., Owen-Smith, J. (2005) Network Dynamics and Field Evolution: The Growth of Interorganizational Collaboration in the Life Sciences, in: *American Journal of Sociology*, 110/4.	92
RWP2005			Robins, G./Pattison, P./Woolcock, J. (2005): Small and Other Worlds: Global Network Structures from Local Processes, *American Journal of Sociology* , 110/4: 894–936.	71
US2005			Brian Uzzi, Jarrett Spiro (2005): Collaboration and Creativity: The Small World Problem, in: *American Journal of Sociology,* 111/2:447ff.	77

Appendix B: Co-cited items[4]

Code	2005	2000	90-92	Citation	Total Co-Citations
AB2002	2			Albert, R., and A.-L. Barabási. 2002. "Statistical Mechanics of Complex Networks." *Review of Modern Physics* 74:4797.	2
Al1979			3	Allan, Graham. 1979. *A Sociology of Friendship and Kinship*. London: Allen & Unwin.	3
ASBS2000	3			Amaral, L. A. N., A. Scala, M. Barthélémy, and H. E. Stanley. 2000. "Classes of Small-World Networks." *Proceedings of the National Academy of Sciences of the United States of America* 97:1114952.	3
Ax1984			2	Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic.	2
Ba1950			2	Bavelas, Alex. 1950. "Communication Patterns in Task Oriented Groups." *Journal of the Acoustical Society of America* 22: 271-282.	2
BA1999	3			Barabási, A. L., and R. Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286:50912.	3
Ba2002	2			Barabasi, A. L. 2002. Linked: The new science of networks.Cambridge, M.A.: Perseus Publishing.	2
Be1997	2			Bearman, P. 1997. "Generalized Exchange." American Journal of Sociology 102:13831415.	2
BG2000	2			Bala, V., and S. Goyal. 2000. "A non-cooperative model of network formation." Econometrica 68:1181-1229	2
Bl1964	2		3	Blau, P. M. 1964. *Exchange and Power in Social Life*. New York: Wiley.	5
Bl1977			3	Blau, P. M. 1977. Inequality and Heterogeneity: A Primitive Theory of Social Structure. Free Press.	3
Bo1985	3			Bollobas, B. 1985. *Random Graphs*. London: Academic Press.	3
BR2002	2			Busch, M. L., and E. Reinhardt. 2002. "Testing International Trade Law: Empirical Studies of GATT/WTO Dispute Settlement." Pp. 457-81 in The Political Economy of International Trade Law: Essays in Honor of Robert E. Hudec, edited by Daniel L. M. Kennedy and James D. Southwick. Cambridge: Cambridge University Press.	2

4 The list of co-cited references is sorted alphetically in the order of codes for each reference, so that finding the reference in the network diagrams is easier. The codes for each cited article were generated by using the first two letters of the author plus the year of the publication. In caseof multipleauthors, the first letter of each last name plus the year (for the first three authors only).

Code	2005	2000	90-92	Citation	Total Co-Citations
BS1984			2	Blau, P. M., and Joseph Schwartz. 1984. *Crosscutting Social Circles*. Orlando, Fla.: Academic Press.	2
Bu1983			2	Burt, R. S. 1983. Range, Chapter 9 in Burt and Minor (eds.) Applied network analysis: A methodological introduction. Beverly Hills: Sage.	2
Bu1992	5			Burt, R. S. 1992. *Structural Holes*. Cambridge: Cambridge University Press.	5
CDD1980			2	Corcoran, M., Linda Datcher, and Greg Duncan. 1980. “Information and Influence Networks in Labor Markets," 1-37, in Duncan, Greg J., and James N. Morgan,eds., Five Thousand American Families, Vol. VIII, Institute for Social Research,University of Michigan.	2
CMH1986			2	Campbell, K., Peter Marsden, and Jeanne Hurlbert. 1986. “Social Resources and Socioeconomic Status.” *Social Networks* 8: 97-117.	2
Co1988		2		Coleman, J. S. 1988. "Social Capital in the Creation of Human Capital." American Journal of Sociology 94(supp.):S95-5120.	2
DG1997	2			Davis, G. F., and Henrich R. Greve. 1997. "Corporate Elite Networks and Governance Changes in the 1980s." American Journal of Sociology 103:1-37.	2
Di1991	2			DiMaggio, P. J. 1991. "Constructing an Organizational Field as a Professional Project: U.S. Art Museums, 1920-1940." Pp. 267-92 in The New Institutionalism in Organizational Analysis, edited by W. Powell and P. J. DiMaggio. Chicago: University of Chicago Press.	2
Dr1998	2			Drentea, Patricia. 1998. "Consequences of Women's Formal and Informal Job Search Methods for Employment in Female-Dominated Jobs." Gender and Society 12: 321-38.	2
DS1988			2	Davis, J. A, and Smith, TW (1988) General Social Surveys, 1972-1988	2
DYB2003	2			Davis, G. F, Mina Yoo, and Wayne Baker. 2003. "The Small World of the American Corporate Elite, 1982-2001." Strategic Organization 3:301-26.	2
Ec1974			2	Eckhoff, T. (1974). Justice: Its determinants in social interaction. Rotterdam, The Netherlands: Rotterdam University Press.	2
Em1962			2	Emerson, R. E., 1962 "Power-Dependence Relations", American Sociological Review, 27:31-40	2
Em1976			2	Emerson, R. M. 1976. “Social Exchange Theory.” *Annual Review of Sociology* 2: 335-62.	2

Code	2005	2000	90-92	Citation	Total Co-Citations
En1992		2		England, P. 1992. *Comparable Worth: Theories and Evidence*. Hawthorne, N.Y.: Aldine de Gruyter.	2
ER1959	2			Erdös, P., and A. Renyi. 1959. "On Random Graphs. I." *Publicationes Mathematicae (Debrecen)* 6:29097.	2
Fe1981			3	Feld, Scott L. 1981. “The Focused Organization of Social Ties.” *American Journal of Sociology* 86: 1015-35.	3
Fi1982a			4	Fischer, C. S. 1982a. To dwell among friends: personal networks in town and city. University of Chicago Press, Chicago.	4
Fi1982b			2	Fischer, C. S. 1982b. “What Do We Mean by ‘Friend’?” *Social Networks* 3: 287-306.	2
Fr1979			3	Freeman, L. C. 1979. “Centrality in Social Networks: Conceptual Clarification.” *Social Networks* 1: 215-39.	3
Fr1980			2	Freeman, L. C., Douglas Roeder, and Robert R. Mulholland. 1980. “Centrality in Social Networks: II. Experimental Results.” *Social Networks* 2: 119-41.	2
Fr1982			2	Friedkin, N. E. 1982. “Information Flow through Strong and Weak Ties in Intraorganizational Social Networks.” *Social Networks* 3: 273-85.	2
FR1991	2			Fruchterman, T. J. J., and Edward Reingold. 1991. "Graph Drawing by Force-Directed Placement." *SoftwarePractice and Experience* 21:112964.	2
FW1997	2	3		Fernandez, R. M., and Nancy Weinberg. 1997. "Sifting and Sorting: Personal Contacts and Hiring in a Retail Bank." American Sociological Review 62:883-902.	5
Ga1995	2			Garebian, K. 1995. The Making of West Side Story. Ontario: Mosaic.	2
Go1960			3	Gouldner, A. W. 1960. “The Norm of Reciprocity: A Preliminary Statement.” *American Sociological Review* 25: 161-78.	3
Go1967	2			Goffman, E. 1967. *Interaction Ritual: Essays on Face-to-Face Behavior*. Garden City, N.Y.: Anchor Books.	2
Go1981	2			Goffman, E. 1981. *Forms of Talk*. Philadelphia: University of Pennsylvania Press.	2
Go2002	2			Gould, R. 1995. *Insurgent Identities: Class, Community, and Protest in Paris from 1848 to the Commune*. Chicago: University of Chicago Press.	2

Code	2005	2000	90-92	Citation	Total Co-Citations
Gr1973	6	2	4	Granovetter, M. S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78	12
Gr1974	3	2	3	Granovetter, M. S., Getting a job: A Study of Contacts and Careers. Cambridge, Mass.: Harvard University Press, 1974. *(Including 1995 2nd edition)*	8
Gr1981		2		Granovetter, M. S., 1981. "Toward a Sociological Theory of Income Differences." Pp.1354 Networks and Employment 11 – 48 in Sociological Perspectives on Labor Markets, edited by Ivar Berg. New York: Academic Press.	2
Gr1982			3	Granovetter, M. S., 1982. "The Strength of Weak Ties: A Network Theory Revisited." Pp. 105-30 in *Social Structure and Network Analysis*, edited by Peter Marsden and Nan Lin. Beverly Hills, Calif.: Sage.	3
Gr1985	5		3	Granovetter, M. S. (1985): Economic Action and Social Structure: The Problem of Embeddedness, in: American Journal of Sociology, 91: 481-510.	8
Ha1983		2		Hammer, M. 1983. "`Core' and `Extended' Social Networks in Relation to Health and Illness." Social Science and Medicine 17:40511.	2
Ha1988			2	Halaby, C. N. (1988). "Action and Information in the Job Mobility Process: The Search Decision." American Sociological Review 53: 9–25	2
He1979	2			Heckman, J. J. 1979. "Sample Selection Bias as a Specification Error." Econometrica 47:153-61.	2
He1994		2		Hedstroem, P. 1994. "Contagious Collectivities: On the Spatial Diffusion of Swedish Trade Unions, 1890-1940." *American Journal of Sociology* 99: 1157-79.	2
HL1971	2			Holland, P. W., and S. Leinhardt. . 1971. "Transitivity in Structural Models of Small Groups." *Comparative Group Studies* 2:10724.	2
Ho1950	2			Homans, G. C. 1950. The Human Croup. Cambridge, Mass.: Harvard University Press.	2
Ho1974			4	Homans, G. C. 1974. *Social Behavior: Its Elementary Forms*, 2d ed. New York: Harcourt Brace Jovanovich.	4
HP1991	2			Hanson, S., and Geraldine Pratt. 1991. "Job Search and the Occupational Segregation of Women." Annals of the Association of American Geographers 81 (2): 229-53.Hardin, Russell. 2002. Trust and Trustworthiness. New York: Russell Sage.	2

Code	2005	2000	90-92	Citation	Total Co-Citations
IR2000	2			Ingram, P., and Peter Roberts. 2000. "Friendships among Competitors in the Sydney Hotel Industry." American Journal of Sociology 106:387-423.	2
JW2002	2			Jackson, M. O., and A. Watts. 2002. "The Evolution of Social and Economic Networks." Journal of Economic Theory 106:265-295.	2
KB1983			2	Knoke, D., and Ronald S. Burt. 1983. "Prominence." Pp. 195-222 in *Applied Network Analysis: A Methodological Introduction*, edited by R.S. Burt and M.J. Minor, Beverly Hills, Calif.: Sage.	2
KK1989	2			Kamada, T., and Satoru Kawai. 1989. "An Algorithm for Drawing General Undirected Graphs." *Information Processing Letters* 31:715.	2
Ko1994	2			Kollock, P. 1994. "The Emergence of Exchange Structures: An Experimental Study of Uncertainty, Commitment, and Trust." American Journal of Sociology 100 (2): 313-34.	2
Kr1987	2			Krackhardt, D. 1987. "QAP Partialling as a Test of Spuriousness." Social Networks 9:171–86.	2
Kr1988	2			Krackhardt, D. 1988. "Predicting with Networks Nonparametric Multiple Regression Analysis of Dyadic Data", Social Networks, 10: 359-81.	2
KW2001	2			Kogut, B., and Gordon Walker. 2001. "The Small World of German Corporate Networks in the Global Economy." American Sociological Review 66:317-35.	2
La2001	2			Lazer, D. 2001. "The Co-evolution of Individual and Network." Journal of Mathematical Sociology 25 (1): 69-108.	2
LD1992		2		Land, K.C., and G. Deane. 1992. "On the Large-Sample Estimation of Regression Models with Spatial or Network Effects Terms: A Two-Stage Least-Squares Approach." Pp. 221-48 in *Sociological Methodology* 1992, edited by P. Marsden. Oxford: Basil Blackwell.	2
Le1951			2	Leavitt, H. J. 1951. "Some Effects of Certain Communication Patterns on Group Performance." *Journal of Abnormal and Social Psychology* 46: 38-50.	2
Le1960		2		Le Bon, G. (1895) 1960. *The Crowd*. New York: Viking	2
Le1988	2			Leifer, E. M. 1988. "Interaction Preludes to Role Setting: Exploratory Local Action." *American Sociological Review* 53:86578.	2

Code	2005	2000	90-92	Citation	Total Co-Citations
Le1993	2			Leidner, R. 1993. Fast Food, Fast Talk: Service Work and the Routinization of Everyday Life. Berkeley and Los Angeles: University of California Press.	2
LEV1981	2	2	2	Lin, N., Walter M. Ensel, and John C. Vaughn. 1981. "Social Resources and Strength of Ties: Structural Factors in Occupational Status Attainment." American Sociological Review 46 (4): 393-405.	6
Li1982			3	Lin, N. 1982. "Social Resources and Instrumental Action." Pp. 131-45 in Social Structure and Network Analysis, edited by Peter V. Marsden and Nan Lin. Beverly Hills, CA : Sage.	3
LP1976			2	Laumann, E. O., and Franz U. Pappi. 1976. *Networks of Collective Action*. New York: Academic Press.	2
LP1999	2			Lazega, E., and P. E. Pattison. 1999. "Social Capital, Multiplex Generalized Exchange and Cooperation in Organizations: A Case Study." *Social Networks* 21:6790.	2
LY1998	2			Lawler, E., and Jeongkoo Yoon. 1998. "Network Structure and Emotion in Exchange Relations." American Sociological Review 63 (6): 871-94.	2
Ma1983			3	Marsden, P. V. 1983. "Restricted Access in Networks and Models of Power." *American Journal of Sociology* 88: 686-717.	3
Ma1987	2		2	Marsden, P. V. 1987. "Core Discussion Networks of Americans." American Sociological Review 52:122-31.	4
Ma1988			2	Masrden, P. V. 1988. "Homogeneity in Confiding Relations." *Social Networks* 10: 57-76.	2
Ma1991	2			Maynard, D. W. 1991. "Interaction and Asymmetry in Clinical Discourse." *American Journal of Sociology* 97:448–95.	2
MB2003	2			McFarland, D., and Skye Bender-deMoll. 2003. "Classroom Structuration: How Interaction Patterns Get Reproduced and Transformed." Working Paper. Stanford University.	2
MC1984			2	Marsden, P. V., and Karen E. Campbell. 1984. "Measuring Tie Strength." Social Forces 63: 482-501.	2
Mc1988		2		McAdam, D. 1988. *Freedom Summer*. New York: Oxford University Press.	2
Mc2001	2			McFarland, D. 2001. "Student Resistance: How Formal and Informal Organization of Classrooms Facilitates Student Defiance." *American Journal of Sociology* 107 (3): 61278.	2

Code	2005	2000	90-92	Citation	Total Co-Citations
Me1968		2		Merton, R. K. 1968. Social Theory and Social Structure. 3d ed. New York: Free Press.	2
Me1973	2			Merton, R. K. 1973. The Sociology of Science. Chicago: University of Chicago Press.	2
MH1988	2		2	Marsden, P. V. and Jeanne S. Hurlbert. 1988. "Social Resources and. Mobility Outcomes: A Replication and Extension." Social Forces. 66:1038-59.	4
Mi1967	5			Milgram, S. 1967. "The Small World Problem." *Psychology Today* 2:6067.	5
MMS1997		2		Munch, A., J. Miller McPherson, and Lynn Smith-Lovin. 1997. "Gender, Children, and Social Contact: The Effects of Childrearing for Men and Women." American Sociological Review 62:509-20.	2
Mo1991		2		Montgomery, J. D. 1991. "Social Networks and Labor-Market Outcomes: Toward an Economic Analysis." American Economic Review 81:1408-18.	2
Mo2004	2			Moody, J. 2004. "The Structure of a Social Science Collaboration Network: Disciplinary Cohesion from 1963-1999." American Sociological Review 69:213-38.	2
MP1990		2		Marsden, P. V., and J. Podolny. 1990. "Dynamic Analysis of Network Diffusion Processes." Pp. 197-214 in *Social Networks through Time*, edited by J. Weesie and H. Flap. ISOR: University of Utrecht.	2
MSC2001	2			McPherson, M., Lynn Smith-Lovin, and James Cook. 2001. "Birds of a Feather: Homophily in Social Networks." Annual Review of Sociology 27:415-44.	2
MW2003	2			Moody, J., and Douglas R. White. 2003. "Social Cohesion and Embeddedness: A Hierarchical Conception of Social Groups." American Sociological Review 68: 103-27.	2
MWP1988			3	Markovsky, B., David Willer, and Travis Patton. 1988. "Power Relations in Exchange Networks." *American Sociological Review* 53: 220-36.	3
Ne2000	2			Newman, M. E. 2000. "Models of the Small-World: A Review." *Journal of Statistical Physics* 101:81941.	2
Ne2001	3			Newman, M. E. 2001 "The Structure of Scientific Collaboration Networks." Proceedings of the National Academy of Sciences 98:404-9.	3
Ne2003	2			Newman, M. E. 2003. "The Structure and Function of Complex Networks." *SIAM Review* 45:167256.	2

Code	2005	2000	90-92	Citation	Total Co-Citations
NSW2001	2			Newman, M. E., Steve Strogatz, and Duncan Watts. 2001. "Random Graphs with Arbitrary Degree Distributions and Their Applications." Physical Review E 64:1-17.	2
PA1993	3			Padgett, J. F., and Christopher K. Ansell. 1993. "Robust Action and the Rise of the Medici, 14001434." *American Journal of Sociology* 98:12591319.	3
Pe1992		2		Pescosolido, B. A. 1992. "Beyond Rational Choice: The Social Dynamics of How People Seek Help." American Journal of Sociology 97:1096-1138.	2
Po1993	2			Podolny, J. M. 1993. "Status-Based Model of Market Competition." American Journal of Sociology 98 (4): 829-72.	2
PS1978			2	Pfeffer, J. and G. R. Salancik, 'The External Control of Organizations: A Resource Dependence Perspective' (Harper & Row, 1978).	2
PSH1996	2			Podolny, J. M., Toby E. Stuart, and Michael T. Hannan. 1996. “Networks, Knowledge, and Niches: Competition in the Worldwise Semiconductor Industry, 1984–1991.” American Journal of Sociology 102:659–89.	2
PSS2000	2			Petersen, T., Ishak Saporta, and Marc-David L. Seidel. 2000. "Offering a Job: Meritocracy and Social Networks." American Journal of Sociology 106 (3): 763-816.	2
Ro1985			2	Rosenthal, C. 1985. “Kinkeeping in the Familial Division of labor.” *Journal of Marriage and the Family* 47 (November): 965-74.	2
Sa1995	2			Salancik, G. 1995. "Wanted: A Good Network Theory of Organization." Administrative Science Quarterly 40:345-49.	2
Sp1974	2	2		Spence, A. M. 1974. Market Signaling: Informational Transfer in Hiring and Related Processes. Cambridge, Mass.: Harvard University Press.	4
St1990		2		Stacey, J. 1990. Brave New Families. New York: Basic Books.	2
St2001	2			Strogatz, S. H. 2001. "Exploring Complex Networks." *Nature* 410:26876.	2
Wa1999	4			Watts, D. J. 1999. Small Worlds: The Dynamics of Networks between Order and Randomness. Princeton, N.J.: Princeton University Press.	4
Wa1999	2			Watts, D. J. 1999. “Networks, Dynamics, and the Small-World Phenomenon.” *American Journal of Sociology* 105:493–527.	2

Code	2005	2000	90-92	Citation	Total Co-Citations
We1982			2	Wegner, B. 1982 (ed) Social Attitudes and Psychophysical measurement. Hillsdale, NJ. Erlbaum.	2
We1983			2	Wellman, B. 1983. "Network Analysis: Some Basic Principles." Pp. 155-200 in *Sociological Theory* 1983, edited by R. Collins. San Francisco: Jossey-Bass.	2
We1985			2	Wellman, B. 1985. "Domestic Work, Paid Work and Net Work." Pp. 159-91 in *Understanding Personal Relationships*, edited by Steve Duck and Daniel Perlan, London: Sage.	2
We1988			2	Wellman, B. 1988. "The Community Question Re-evaluated." Pp. 81-107 in *Power, Community and the City*, edited by Michael Peter Smith. New Brunswick, N.J.: Transaction.	2
WF1994	4			Wasserman, S., and Katherine Faust. 1994. *Social Network Analysis*. Cambridge: Cambridge University Press.	4
Wh1970	2			White, H. C. 1970. "Search Parameters for the Small World Problem." Social Forces 49:259-64.	2
Wh1981	2			White, H. C. 1981. "Where do markets come from?" American Journal of Sociology 81: 730-79.	2
Wi1985			2	Williamson, O. E. 1985. *The Economic Institutions of Capitalism*. New York: Free Press.	2
WS1998	4			Watts, D. J., and Steven H. Strogatz. 1998. "Collective Dynamics of `Small-World' Networks." Nature 393:44042.	4
YGC1988			3	Yamagishi, T., Mary R. Gillmore, and Karen S. Cook. 1988. "Network Connections and the Distribution of Power in Exchange Networks." American Journal of Sociology 93: 833-851.	3