In the Proceedings of the Annual Conference of Association for Library and Information Science Education (ALISE)
January 20-23, 2009
Denver, CO

Name Networks: A Content-Based Method for Automated Discovery of Social Networks to Study Collaborative Learning

Anatoliy Gruzd

Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign
agruzd2@illinois.edu

Abstract

As a way to gain greater insight into the operation of Library and Information Science (LIS) e-learning communities, the presented work applies automated text mining techniques to text-based communication to identify, describe and evaluate underlying social networks within such communities. The main thrust of the study is to find a way to use computers to automatically discover social ties that form between students just from their threaded discussions. Currently, one of the most common but time consuming methods for discovering social ties between people is to ask questions about their perceived social ties via a survey. However, such a survey is difficult to collect due to the high cost associated with data collection and the sensitive nature of the types of questions that must be asked. To overcome these limitations, the paper presents a new, content-based method for automated discovery of social networks from threaded discussions dubbed *name networks*. When fully developed, name networks can be used as a real time diagnostic tool for educators to evaluate and improve teaching models and to identify students who might need additional help or students who may provide such help to others.

1 The Use of Social Network Analysis in E-Learning Assessment

Social Network Analysis (SNA) is a common method to study social interactions and collaborative learning of online groups. Some examples of studies that relied on SNA to evaluate individual learning based on the position of individuals in a network and group cohesion based on general properties of a network include Cho et al. (2007), Reyes & Tchounikine (2005) and Willging (2005). From the social network perspective, individual behavior is defined by others. Thus, to understand individual behavior, we need to "describe patterns of relationships between actors, analyze the structure of these patterns, and seek to uncover their effect on individual behavior" (Nurmela et al., 1999, n.p.). In any social network, there are *nodes* which represent group members, and edges (often referred to as *ties*) that connect people by means of various types of relations. The strength of the relations is usually conveyed via a *weight* assigned to each *tie*.

A traditional way to collect information about social networks within communities is to ask group members themselves via a survey. However, this method is very time consuming and prone to a high rate of non-responses. Dillman (2000) posited two main reasons for the high rate of non-responses: the questions are highly personal, and the procedures for answering the questions are too burdensome. As a result of these inherent flaws with surveys data, many researchers are focusing on cheaper and more objective, automated methods for collecting data on social networks. Some of these automated methods include using movement tracking devices (e.g., Matsuo et al., 2006), log analysis (e.g., Nurmela et al., 1999), and co-citation analysis (e.g., White et al., 2004).

The most common automated method to collect information on social networks in online communities is to gather 'who replies to whom' data which counts the number of messages exchanged between individuals from their recorded interactions. Higher number of messages exchanged is usually interpreted as stronger ties between people. This method is often used with email-types data. In online communities that use threaded discussions, researchers usually rely on information in posting headers about the chain of people who had previously posted to the thread (further referred to as *reference chain*) to gather 'who replies to whom' data. For logical and practical reasons, in the past researchers generally assumed that the reference chain may reveal the addressee(s). More specifically, it is usually assumed that a poster is replying to the previous poster in the reference chain. (For the remainder of this discussion, I will refer to any social network that is built using information in the reference chain as a *chain network*.) Unfortunately, this assumption is not always 100% true in highly argumentative and collaborative communities such as online classes. A previous poster is not always an addressee of the posting and vice versa. A poster may address or reference other posters from earlier in the thread, from another thread, or even from other channels of communication (e.g., emails, chats, face to face meetings, etc). So, while the use of reference chains provides some mechanism to approximate 'who replies to whom' data for threaded discussions, such approximation is not very accurate and is likely to cause an undercounting of possible connections. To overcome the inherent flaws associated with gathering 'who replies to whom' data from threaded discussions, I propose a hybrid approach called *name networks* for inferring social networks using both the posting headers and the actual content of postings. The next section will briefly describe the procedure for building name networks.

2 Building Name Networks

In general, to build name networks, the method starts by automatically finding all mentions of personal names in the postings and uses them as nodes in the name networks. Next the method proceeds to discover ties between all the nodes by connecting a poster to all names found in his/her postings. Finally, to disambiguate name aliases, the algorithm adopts a simple but effective approach that relies on associating names in the postings with email addresses in the corresponding posting headers. For more detailed description of the method, see Gruzd & Haythornthwaite (2008b).

Personal names were chosen as the main input into building name networks because they have been shown to be good indicators of social ties. Linguistically speaking, the use of personal names performs two main communicative functions as identified by Leech (1999): (1) addressee-identifying and attention-getting, (2) social bond-maintaining function. The first function is self-explanatory, when calling somebody by his/her name, a person identifies somebody among others to talk to and at the same time tries to get that person's attention. As for the social bond-maintaining function, its main purpose is to maintain and reinforce social relationships. For example, when someone uses formal names and titles, it might be to indicate subordination in the relationships. While at the opposite end, someone might use informal names or nicknames to show the same social status or emphasize friendship. The social bond-maintaining function of naming is especially important in online groups. Since names are "one of the few textual carriers of identity" in discussions on the web (Doherty, 2004, p. 3), their use is crucial to create and maintain a sense of community (Ubon, 2005) and social presence (Rourke, 2001). As Ubon (2005) put it, by addressing each other by name, participates "build and sustain a sense of belonging and commitment to the community" (p.122). In sum, by focusing on personal names, the "name network" method can quickly identify addressees of each message and thus automatically discover "who talks to whom" in one-to-many types of online communication such as threaded discussions and chats.

Furthermore, the social bond-maintaining function of personal names suggests that the discovered ties between people will not just reflect communication patterns, but also likely reflect real social relationships between people.

To evaluate the proposed method of building social networks and identify what exactly will be gained from using this new method, social networks derived using the 'name network' method will be compared against those derived from other means, specifically chain (reply-to) networks and students' self-reported (perceived) social networks. For the purpose of this work, chain networks will be built by connecting a sender to the most recent poster in the thread, while self-reported social networks will be built based on the data collected via an online questionnaire.

3 Data Collection and Self-Reported Networks

The dataset for this study consists of bulletin board postings and students' responses to an online questionnaire from six different graduate level online classes at the Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. The data was collected in Spring 2008 as part of a larger study on online learning in collaboration with Caroline Haythornthwaite. Prior to the beginning of the data collection, Institutional Review Board permission was obtained for this work. All students' names in the study were ammonized to protect their privacy.

Instructors in these classes primarily relied on Moodle (an open source course management system) to make announcements, distribute class materials and facilitate weekly discussions using bulletin boards. Once a week, students met online using LEEP, a home-built online environment. During these live sessions, the instructor delivered the lecture via a live audio feed. During the lecture, students could ask questions or answer instructor's questions by typing in the chat room. During some live sessions, the instructor divided students into smaller discussion groups (each group would use a separate chat room for their discussions).

Students' self-reported social networks were collected via an online questionnaire administered at the end of the semester. The first group of questions asked students to indicate the frequency of their associations with each classmate on a scale from 1 to 5 (with [5] indicating a more frequent association) with respect to three different relations: learning something new about the subject matter from another student, working together, and friendship. The second group of questions asked students to nominate 5 to 8 prominent students that best fit the following four criteria: "influential in one's learning", "important in promoting discussion", "help with understanding a topic or assignment" and "made class fun". The response rate for the questionnaire was 64% (a total of 81 responses). Each question was designed to discover one of the many possible social relations (e.g., learn, work, help, etc) that might exist between the students.

A self-reported network was then built using the following procedure. First, the procedure added a tie between each respondent and his or her nominees. For the questions from the first group, only nominations with an association level of 3 or higher were considered. The next step was to assign weights to each tie. The weights were assigned based on how many times each nominee was selected by the same respondent. To better reflect actual social relationships between students, the procedure removed all "weak" ties with a weight of less than 3. Since the procedure only kept so-called "strong" ties, it is very likely that they will be symmetric. To help restore some ties missing due to the non-respondents, the resulting network was symmetrized.

Open source software called phpESP (http://phpesp.sourceforge.net) was used to conduct the survey. A Social Network Analysis tool called ORA v.1.9.5.2.6 (http://www.casos.cs.cmu.edu/projects/ora) was used for storage and basic manipulations of the network data. Internet Community Text Analyzer (ICTA) (http://textanalytics.net) described in Gruzd & Haythornthwaite (2008a) was used to automatically build name and chain networks.

4 Chain Networks versus Name Networks

First, the analysis began with comparing name and chain networks using QAP correlations (Krackhardt, 1987). This was done to determine the level of overlap between these two types of networks. QAP correlation relies on Pearson's correlation coefficient to compare relational data. It was chosen as the method of measurement for this work because "it presumes neither random sampling of cases from a population […] nor independence of observations" (White et al., 2004, p. 116).

Software called ORA was used to compute the QAP correlations. The results of the comparison are presented in Table 1 below. All tests were significant ($p<=0.05$). In all classes, pairs of name and chain networks demonstrated moderate correlations between 0.45 and 0.69 (See the "QAP" column in Table 1). As expected, there is some overlap between posting behavior as represented by the chain networks and "naming" behavior as represented by the name networks. However, there are also substantial differences in what is revealed by each of these networks. To better understand these differences and assess the accuracy of chain networks, the next section will compare all connections that make up each tie from the name network with those from the chain networks. More specifically, the next step in the analysis will determine how many connections discovered by the 'name network' method were not discovered by the 'chain network' method.

Table 1. QAP correlations between pairs of the name and chain networks for six online classes

	# of Students	Chain Network Density	Name Network Density	QAP correlations*
Class1	28	0.23	0.13	0.50
Class2	20	0.48	0.35	0.51
Class3	25	0.48	0.28	0.58
Class4	21	0.08	0.1	0.45
Class5	19	0.22	0.15	0.53
Class6	15	0.39	0.17	0.69

* The number of random permutations used for the analysis was 5,000

Chain networks are built based on the information in the reference chain; as a result, they will fail to connect a poster to poster's addressee whose email is not yet in the reference chain. This situation can arise in one of two ways: (1) when it is a first posting of a new thread or (2) when an addressee has not posted anything to an existing thread. Since all of the names extracted for building name networks were manually inspected for accuracy in the study, it is fair to use these names as actual addressees of postings or people who are somehow connected to the poster. Using an automated script, I counted the number of instances for each of the two situations described above. The counts revealed some pleasantly unexpected results (See Table 2). On average, a chain network misses about 33% of the potentially important connections as compared to the name network. Of the 33% missed connections, 23% came from postings that were the thread starters (Column A) and about 10% came from subsequent messages in a particular thread (Column B). Additionally, there were another 7% of missed connections that occurred when an actual addressee or a 'reference' person was the author of a previous posting in the thread, but not the most recent one (Column D).

Table 2. The relationship between an actual addressee and his/her position in the reference chain

Class	# of all postings*	# of found instances of named addressees	# of times an addressee is <u>NOT</u> in the reference chain when found in …		# of times when an addressee is <u>IN</u> the reference chain as …	
			a first posting of a new thread	a subsequent posting in a thread	the most recent poster	other
			COLUMN A	COLUMN B	COLUMN C	COLUMN D
Class1	608	149	50	11	81	7
Class2	855	271	59	30	153	29
Class3	1,502	306	37	21	232	16
Class4	164	96	17	16	51	12
Class5	412	156	46	26	76	8
Class6	497	107	27	4	73	3
Average (%)		100%	**23%**	**10%**	60%	**7%**

* On average, about 25% of all postings included personal names

To determine the exact nature of connections that were missed by the chain networks, I analyzed all postings that correspond to columns A and B in Table 2 for all six classes.

Situation 1: First Posting of a Thread. The semi-automated content analysis of postings using ICTA revealed that among the most commonly used names in the first posting of a new thread was the instructor's name. Specifically, instructor's name was used to

- Ask the instructor about something (e.g., "*[Instructor's name] if you see this posting would you please clarify for us*"),
- Ask peers to clarify something that the instructor said during the lectures (e.g., "*I remember* [Instructor's name] *asking us to email her with topics [...] I wonder if that is in replacement of our bb question?*"), or
- Share information with classmates obtained from the instructor via some other personal communication such as email. (e.g., *"I just got a reply from [Instructor's name], and she said that […]*")

This type of postings and the ties derived from them is very important in the context of learning. This is because "student-instructor" ties derived from these messages can be used to identify students who are repeatedly asking for instructor's help. For example, a high weight for a tie between a student and the instructor may suggest that a student is uncertain about something in the class and might need extra attention from the instructor. However, if many students are connected to the instructor via these types of messages, then it may indicate that lectures or other class materials are unclear to not just one student and thus either the materials or a delivery method might need to be reconsidered by the instructor.

Another common category of messages was when an instructor mentioned a student. These were usually announcements from the instructor containing names of students responsible for leading a class discussion. For example, "*Dan, [...] Since you have studied [Topic], would you get our discussion going on the forum for this week*". Sometimes an instructor would also mention a student praising him or her for some good work in the class. This suggests that if there is a tie from an instructor to a student based on this kind of postings, then it is very likely that this student is doing well in the class. Identifying reliable and successful students in a class is an important task for any instructor or school's administration, especially when formal grading information is unavailable. For example, an instructor can use such information to assign students into more effective groups in which at least one of the students is doing above average in the class.

Another common type of messages in this category was when an instructor listed groups with their individual members for smaller group discussions. After examining these postings, I concluded that the ties derived from them do not necessarily reflect relationships between the instructor and a student. Instead, these postings can be used to automatically identify students who were assigned to work together, thus potentially creating “work” ties. “Work” ties are especially important for studying online groups since they are often precursors of even closer ties between online participates (See, for example, Haythornthwaite, 2002). This was confirmed by several students in the comment section of the online survey. They viewed the break down into smaller groups during live sessions as a good way to get to know their peers.

The last category of messages was when a student mentioned other student(s). In these cases, the poster often took a leadership role in a group, for example, by summarizing other group members’ postings or assigning roles for a project as demonstrated in the following excerpt:

> “*Some quick poking around shows that Steve and myself are here in Champaign, [...] and Nicole is in Chicago. [...] does anyone have a strong desire to be our contact person to the administrators*”

This type of messages is useful in identifying active group members and group leaders and would be very useful when studying collaborative learning. However, a lot of messages like this from the same person may be perceived negatively by other group members. For example, in a related study, when analyzing a large collection of Usenet newsgroup messages, Fiore et al. (2002) found that online participates who dominated the conversations were often viewed unfavorably. Nevertheless, a more detailed analysis is needed to study the influence of this type of messages/connections in the online learning environment.

Situation 2: Subsequent Posting in a Thread. The detail examination of this type of messages revealed three main types of references/connections:

- A reference to an event or interaction that happened outside the bulleting board (e.g., “*Dan and I have been corresponding via e-mail and he reminded me that we should be having discussion here*"). This type of messages is likely to connect people who work together. It is also suggestive of stronger personal ties. This is because according to the idea of media multiplexity, stronger ties tend to communicate via more communication channels (See, for example, Haythornthwaite & Wellman, 1998; Haythornthwaite, 2001).
- A reference to someone as part of a group when providing a feedback to the whole group or posting on behalf of the whole group and signing the names of all group members (e.g., “*Angela and Natasha, I couldn't wait to see your site. I knew it was going to [be] awesome!*”). This is another type of messages that will likely indicate “work”-related ties.
- A reference to somebody who presented or posted something awhile ago or via different communication channel (e.g., “*[…] it made me think of the faceted catalogs' display that Susan posted*”). These postings are likely to identify “learning” ties. This is because they show that a poster was not just commenting on the previous post, but rather on something that was said awhile ago. This means that the poster was following the class discussion. And a student mentioned in the posting made some significant contribution to the discussion that resonated with the current poster. All these activities can be categorized as evidence of learning.

5 Chain Networks and Name Networks versus Self-Reported

The final part of the study was to compare chain and name networks with self-reported networks and to determine which of the two networks is a better approximation of self-reported (perceived) social networks (if any). Traditionally it is presumed that observed social networks such as chain networks can more accurately reflect actual relations between group members as opposed to their individually perceived perspectives and thus provide a better representation of what is really going on in an online community. But for online learning environments, due to the individualized nature of learning itself, it may be more important to identify and understand perceived social networks in the context of studying collaborative learning. This is because what is deemed as important or relevant to one student may only be marginally valued by another student. Until now, the only reliable way to collect perceived data has been through resource-demanding surveys. Therefore, it would be a methodological breakthrough if an automated method for mimicking perceived social networks is devised.

For this analysis, I conducted a pair wise comparison of the three types of networks using statistical network models and specifically Exponential Random Graph models or just p* models (Robins, In press). To build p* models, I used XPNET software (Wang et al., 2006). There are a few important reasons why p* models were selected to conduct this comparison and not other statistical models or QAP correlations. First, since some students did not participate in the survey, some possible ties were probably missing in the self-reported networks. As a result, QAP correlations would likely produce inadequately lower results. Second, parameters estimated by p* model is easy to interpret and compare across different pairs of networks. Finally, p* model is the only statistical model that is capable of modeling different network structures as well as individual characteristics of the group members (Snijders, 2008).

Using p* models, for each class I estimated the parameter EdgeAB for a pair of the chain network and self-reported network first and then for a pair of the name network and self-reported network. The parameter EdgeAB indicates the likelihood of two networks sharing ties not by a chance alone. The results are shown in Table 3. The model was converged (t–statistics<0.1 for all estimated parameters) and the model was found to be significant (the goodness of fit for EdgeAB was less than 0.1 and between 1 and 3 for all other parameters) for all classes, except the case of a pair of the name and self-reported networks for Class6.

Table 3. EdgeAB - the likelihood of two networks to share ties not by a chance alone

Class	Chain* & Self-Reported Networks		Name* & Self-Reported Networks	
	Estimated parameter EdgeAB	**t-statistics**	**Estimated parameter EdgeAB**	**t-statistics**
Class1	0.81	0.075	1.73	-0.085
Class2	0.99	0.044	1.52	0.031
Class3	1.17	-0.057	1.31	0.001
Class4	0.61	-0.007	1.11	0.064
Class5	1.03	-0.004	0.96	-0.071
Class6	1.33	0.053	0.82	Not significant

* Because self-reported networks likely include only strong ties (Bernard et al. 1981), all weak ties (with weights less than 2) were removed from all chain and name networks (except those for Class4 due to its low network density). Following the requirements of XPNET, both chain and name networks were then binarized, a process where all weights of existing ties were set to 1. Finally, all networks were symmetrized using the following procedure: if there is a connection between one student to another, then it was assumed that for strong ties there is also a connection in the opposite direction.

The results show that for four out of six classes, the name networks are consistently more likely to share ties with the self-reported networks than the chain networks (more than just by chance alone). This supports my general expectation that the name networks are more reflective of students' perceived relationships. However, for two smaller classes, Class5 and Class6, the name networks were less likely to match the self-reported networks. (For Class6, the model was not significant.) This was a very puzzling but intriguing result. It led to a separate investigation that is currently on the way. Below are some preliminary results that suggest some concrete steps on how to further improve the “name network” method in the future to better reflect self-reported (perceived) social networks.

To find out why the name networks for Class5 and Class6 were less likely to share ties with the self-reported networks than the chain networks, I decided to analyze the network signatures for each student in both classes to discover specific differences between these two types of networks. In the current paper, only preliminary results for two students, Nick and Anna from Class5 are reported below. These two students were selected because their network signatures were the most different in each of the two types of networks. One student, **Nick**, had several ties in the self-reported network that were missing in the name network. The second student, **Anna**, had a couple of ties in the name network that were missing in the self-reported network. For these two students, I examined all of their ties that exist in the self-reported network but not in the name network and vice versa. One of the main goals of this analysis is to identify what caused the “name network” method to miss some self-reported ties and to include some ties that are not in the self-reported network. Furthermore, the analysis will help to identify any additional clues from the content of postings that can be used to improve the “name network” extractor. For this examination, I used ICTA.

5.1 Why did the Class5 name network miss some self-reported ties?

A student named **Nick** from Class5 was selected by seven other students in the self-reported survey, but strangely in the name network, Nick was not connected to any of these seven individuals. After a brief investigation, it was determined that Nick only posted three messages to the bulletin board for the whole semester. There was simply not enough evidence on the bulletin board for the name network to discover ties to other individuals. So, on the surface, it is not clear what was the basis for these 7 nominations from his fellow students. A posting from the instructor can shed some light on this mystery. The instructor mentioned Nick on the bulletin board once, when assigning students into smaller discussion groups for the chat sessions. It turned out that the other two students who were assigned to work with Nick were among those who nominated Nick in the survey. This suggested an important future improvement to the ‘name network’ method. In addition, to connecting a poster with all people who are mentioned in the body of his or her posting, the ‘name network’ method should also connect any people whose names co-occur in close proximity in the same messages. With such a modification, Nick would gain two more additional ties in the name network to the two students who nominated him in the survey. As a proof of the concept, I re-built the name network for this class using co-occurrence of names in the text as an additional indicator of personal ties and re-run the comparison analysis between the name network and self-reported network for Class5. This time the likelihood of sharing ties between these two networks increased to **1.50** (t-statistics=0.067) which is higher than the corresponding value from the chain network.

5.2 Why did the Class5 name network include some ties that were not in the self-reported network?

Anna is a well connected student in the self-reported network. However, she only had three strong ties in the name networks. For the purposes of this section, I will only focus on two of the three ties from the name network that are missing in the self-reported network. (The third tie was reported in the self reported network and thus is not relevant to this part of the discussion.) The two ties in questions are with fellow students **Rick** and **Mark**.

The tie between **Anna** and **Rick** resulted from Rick posting three different messages to Anna thanking her for "*insights*", "*thoughtful comments*", and "*all the wonderful posts and information*". However, surprisingly there was no tie between these two students in the self-reported network. After a detail investigation, it turned out that Rick did select Anna in the survey as a person who influenced his learning and helped the most in the class. (Rick was not nominated by Anna.) But because all ties with a weight less than 3 were removed (See Section 3), a tie of 2 between Anna and Rick also disappeared. As an experiment, I built a "learning" network based on the students' responses to only one of the question in the survey about "learning". In this learning network, there was a tie between Anna and Rick. Next I compared this "learning" network with the original name network (without using co-occurrences). The resulting likelihood has slightly increased from 0.96 to 1.17 (t-statistics = -0.062). This suggests that the name network was a bit more similar to the "learning" network than to the overall self-reported network for this particular class. Therefore, the continuation of this study will be to compare name networks with each type of self-reported networks to determine if name networks are better in predicting "learning" ties than others. However, it is possible that for some other class, depending on the prevalence of one type of interactions over the other, a name network can better reflect other types of self-reported networks such as "friendship" or "work" networks. Therefore, as a future improvement, the "name network" method should be able to not just discover ties but also categorize them into different relations. This can be done by using information about roles of participants (e.g., student, guest speaker, instructor, etc), a position of a message in the thread as suggested in Section 4, and/or the context words where particular names are mentioned in a posting. For example, words like "thank you", "help", "assistance" may indicate that a student helping another student, thus they are connected via the "help" relation. With such an algorithm in hand, it will be possible to build name networks that reflect only "help" relations, only "learning" relations, only friendship or some other relation that is important to members of a certain online community.

The tie between **Anna** and **Mark** resulted from Mark posting two messages with Anna's name in them. The first posting from Mark was a question directed at Anna, "*Anna -- what did you mean by [word] in paragraph 3 of your reply?*" The second message was a thank you message from Mark to Ann for posting an interesting article to the bulletin board. (There were no messages from Ann mentioning Mark's name.) But regardless, this may be enough to suggest a tie between Mark and Anna. Unfortunately, because Mark did not participate in the survey, the self-reported network did not include a tie between them. In such case, a researcher can rely on tools like ICTA to conduct a semi-automated content analysis of messages to make the final decision about the accuracy of the "name network" method.

6 Conclusions

The "name network" method as proposed and evaluated in this work provides one more option for understanding and extracting social networks from online discussion boards. Section 4 demonstrated that name networks provide on average 33-40% more information about social ties in a group as compared to chain networks. This additional information is available because name networks can account for instances when a poster addresses or references somebody who has not previously posted to a particular thread. Furthermore, the results of the study demonstrated that name networks are also very adept at detecting social relations such as work and help which are considered by many researchers to be crucial in shared knowledge construction and community building. Based on the discussion in Section 5, there is an overall tendency of name networks to better reflect self-reported ties than chain networks. These characteristics make name networks a useful real time diagnostic tool for educators to evaluate and improve teaching models, and to identify students who might need additional help or students who may provide such help to others.

Finally, the semi-automated content analysis of postings from Class5 and Class6 classes using ICTA (http://textanalytics.net) suggested two important improvements to the "name network" method to increase the accuracy of tie discovery. The suggested improvements include (1) using names co-occurrence as an indicator of the "work" relation and (2) identifying types of different relations based on the context words used in the postings.

References

Bernard, H.R., P. Killworth, L. Sailer. 1981. Summary of Research on Informant Accuracy in Network Data and the Reverse Small World Problem. *Connections* 4(2): 11-25.

Cho, H., Gay, G., Davidson, B. and Ingraffea, A. (2007). Social Networks, Communication Styles, and Learning Performance in a Cscl Community. *Computers & Education* 49(2): 309-329.

Dillman, D.A. (2000). Mail and Internet Surveys: The Tailored Design Method. New York, Wiley.

Doherty, C. (2004). Naming the Trouble with Default Settings. In the Proceedings of *"SFL Ripples in the 21st Century" Australian Systemic Functional Linguistics Association Conference*, Brisbane, Australia.

Fiore, A.T., Tiernan, S.L. and Smith, M.A. (2002). Observed Behavior and Perceived Value of Authors in Usenet Newsgroups: Bridging the Gap. In the Proceedings of *the SIGCHI conference on Human Factors in Computing Systems: Changing our world, changing ourselves*, Minneapolis, Minnesota, USA, ACM.

Gruzd, A. and Haythornthwaite, C. (2008a). The Analysis of Online Communities Using Interactive Content-Based Social Networks (Extended Abstract). In the Proceedings of *the ASIS&T 2008 conference*, pp. 523-527.

Gruzd, A. and Haythornthwaite, C. (2008b). Automated Discovery and Analysis of Social Networks from Threaded Discussions. *Paper presented at the International Network of Social Network Analysts*.

Haythornthwaite, C. (2002). Strong, Weak, and Latent Ties and the Impact of New Media. *The Information Society* 18(5): 385-401.

Haythornthwaite, C. (2001). Exploring multiplexity: Social network structures in a computer-supported distance learning class. *The Information Society* 17(3): 211-226.

Haythornthwaite, C. & Wellman, B. (1998). Work, friendship and media use for information exchange in a networked organization. *Journal of the American Society for Information Science* 49(12): 1101-1114.

Leech, G. (1999). The Distribution and Function of Vocatives in American and British English Conversation. In H. Hasselggård and S. Oksefjell (Eds.), *Out of Corpora: Studies in Honour of Stig Johansson*, Rodopi.

Matsuo, Y., Hamasaki, M., Nakamura, Y., Nishimura, T., Hasida, K., Takeda, H., Mori, J., Bollegala, D. and Ishizuka, M. (2006). Spinning Multiple Social Networks for Semantic Web. In the Proceedings of *the Twenty-First AAAI Conference on Artificial Intelligence*, Boston, Massachusetts.

Nurmela, K., Lehtinen, E. and Palonen, T. (1999). Evaluating CSCL Log Files by Social Network Analysis. In the Proceedings of *the 1999 Conference on Computer Support* *for Collaborative Learning*, Palo Alto, California, International Society of the Learning Sciences.

Reyes, P. and Tchounikine, P. (2005). Mining Learning Groups' Activities in Forum-Type Tools. In the Proceedings of *the 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: the Next 10 years!*, Taipei, Taiwan, International Society of the Learning Sciences.

Robins, G. (In press). Exponential Random Graph (P*) Models for Social Networks. In R. Myers (Ed.), *Encyclopaedia of Complexity and System Science*, Springer.

Rourke, L., Anderson, T., Garrison, D. R., & Archer, W. (2001). Methodological Issues in the Content Analysis of Computer Conference Transcripts. *International Journal of Artificial Intelligence in Education* 12: 8-22.

Snijders, T.A.B. (2008). Models for Social Networks. *Course Lecture,* University of Oxford.

Ubon, A.N. (2005). Social Presence in Asynchronous Text-Based Online Learning Communities: A Longitudinal Case Study Using Content Analysis. Department of Computer Science, The University of York. Doctor of Philosophy.

Wang, P., Robins, G. and Pattison, P. (2006). PNet: Program for the Estimation and Simulation of P* Exponential Random Graph Models, User Manual. Department of Psychology, University of Melbourne.

White, H.D., Wellman, B. and Nazer, N. (2004). Does Citation Reflect Social Structure?: Longitudinal Evidence from the Globenet Interdisciplinary Research Group. *Journal of the American Society for Information Science and Technology* 55(2): 111-126.

Willging, P.A. (2005). Using Social Network Analysis Techniques to Examine Online Interactions. *US-China Education Review* 2(9).