Studying Collaborative Learning Using Name Networks

Anatoliy Gruzd
School of Information Management, Dalhousie University. 6100 University Avenue, Halifax, NS B3H 3J5 Canada. E-mail: AGruzd@Dal.Ca

This article reports on efforts to gain greater insight into the operation of LIS e-learning communities, by applying automated text-mining techniques to text-based communication to identify, describe and evaluate underlying social networks within such communities. More specifically, the article presents a new, content-based method for automated discovery of social networks from threaded discussions called *name networks.* This new method has been incorporated into an online tool called Internet Community Text Analyzer (ICTA) and is available at http://textanalytics.net. The proposed method is evaluated in the context of six online classes. The results suggest that the name network method is a viable alternative to costly and time-consuming collection of social network data using surveys. The study also demonstrates how name networks can be used to evaluate online classes.

Keywords: E-learning, online communities, social network analysis, text mining

The Use of Social Network Analysis in E-Learning Assessment

Social Network Analysis (SNA) is method often used to study social interactions and collaborative learning of online groups. 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, Gay, Davidson, and Ingraffea (2007), Reyes and Tchounikine (2005), and Willging (2005). From the social network perspective, individual behavior is defined by others. 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, Lehtinen, & Palonen, 1999, p. 434). In any social network, there are *nodes,* which represent group members, and edges (often referred to as *ties*), which connect people by means of various types of relations. The strength of the relations is usually conveyed by a *weight* assigned to each *tie*.

An established way to collect information about social networks within communities is to survey group members. However, this is very time consuming and prone to a high rate of non-responses (Dillman, 2000). As a result, many researchers are developing cheaper, more objective, and automated methods for collecting social network data. Some 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, Wellman, & Nazer, 2004).

The automated method most commonly used 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

 ISSN: 0748-5786

recorded interactions. A comparatively high number of messages exchanges is usually interpreted as representing a strong tie. This method is often used to analyze e-mail exchanges. In online communities that use threaded discussions, researchers usually rely on information in the headers of posts about the chain of people who previously posted to the thread (referred to as a *reference chain*) to gather 'who replies to whom' data. In the past, researchers generally assumed that the reference chain would reveal the addressee(s). More specifically, it was usually assumed that a poster was replying to the previous post 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, in highly argumentative and collaborative communities such as online classes, this assumption does not always prove true. A previous poster is not always an addressee of the post 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 a mechanism to approximate 'who replies to whom' data for threaded discussions, the 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, this article proposes a new approach called *name networks* for inferring social networks based on the actual content of posts. The next section will briefly describe the procedure for building name networks.

Building Name Networks

To build name networks, all mentions of personal names in the posts are retrieved and used as nodes. Ties between the nodes are identified by connecting each poster to all names found in his or her posts. Finally, to disambiguate name aliases, the algorithm adopts a simple but effective approach that relies on associating names in the posts with e-mail addresses in the corresponding post headers. For a more detailed description of the method, see Gruzd and 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, and (2) social bond-maintaining. The first function is self-explanatory. When calling someone by his or her name, a person identifies that person among others to talk to, and at the same time tries to get that person's attention. The main purpose of the social bond-maintaining function is to maintain and reinforce social relationships. For example, when a person uses formal names and titles, it may be to indicate subordination in the relationships. Conversely, a person may use informal names or nicknames to show the same social status or emphasize friendship. 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 many-to-many types of online communication, such as threaded discussions and chats. Furthermore, the social bond-maintaining function of personal names suggests that the ties discovered between people will not just reflect communication patterns, but also likely reflect the social relationships between people.

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

Data Collection and Self-Reported Networks

The dataset for this study consists of bulletin board posts 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. Each class enrolled between 15 and 28 students. The data was collected in Spring 2008 as part of a larger study on online learning in collaboration with Professor Caroline Haythornthwaite. Prior to the beginning of the data collection, Institutional Review Board permission was obtained for the use of human subjects. All students' names were made anonymous to protect their privacy.

Instructors of the classes primarily relied on Moodle (an open source course management system) to make announcements, distribute class materials and facilitate weekly discussions using online bulletin boards. Once a week, students met synchronously using online software developed by the institution. During these live sessions, the instructors delivered lectures using 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 separate chat rooms for small group discussions.

Students' self-reported social networks were collected through 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 63% (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, a tie was established 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 was very likely that they would be symmetric. To help restore some missing due to the non-respondents, the resulting network was symmetrized.

The open source software application phpESP (http://phpesp.sourceforge.net) was used to conduct the survey. A Social Network Analysis tool, ORA v.1.9.5.2.6 (http://www.casos.cs.cmu.edu/projects/ora), was used to store and manipulate the network data. Internet Community Text Analyzer (ICTA) (http://textanalytics.

net) described in Gruzd and Haythornthwaite (2008a) was used to automatically build name and chain networks.

Name Networks versus Chain Networks

To better understand differences between name and chain networks, this section compares all connections that make up each tie from the name network with those from the chain network. More specifically, the comparison determines how many connections discovered by the name network method are not discovered by the chain network method and vice versa.

Connections Missed by the Chain Network

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 when the addressee's e-mail address is not yet in the reference chain. This situation can arise in one of two ways: (1) when it is a first post in 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, it is fair to use these names as actual addressees of posts or people who are somehow connected to the poster. Using an automated script, the number of instances was counted for each of the two situations described above. The counts revealed some unexpected results (See Table 1). On average, the chain network method misses about 33% of the potentially important connections as compared to the name network method. Of the missed connections, about 70% (or 23% of the total count) came from thread starter posts (Column A), and about 30% (or 10% of the total count) came from subsequent messages in a particular thread (Column B). What stands out about this finding is that the majority of missed addressees (70%) were found in the thread starting messages. This discovery is especially interesting given that most previous research on interactivity in online discussions usually relied on 'reply' messages to estimate class interactivity (e.g., Bonnett et al., 2006). However, these findings suggest that thread starting posts should not be ignored and can also be very interactive in nature. This is because many of these types of posts tend to include references to discussions that occurred among the community members in different threads. Since the name network method is capable of capturing connections to other group members even in the thread starting messages, the method also may be a good model for studies on group interactivity. Future research is needed to confirm this observation.

Another 7% of connections that were missed by the chain network method (Column D) were connections that occurred when an actual addressee or a 'reference' person was the author of a previous post in the thread, but not the most recent one. This happened because the chain network method connects a message sender to the most recent poster in the thread. Generally speaking, it is easy to revise the chain network method to include these 7% of missed connections. The revised method would connect a poster to all previous posters or to two most recent posters in the thread. However, this will likely also introduce more false-positive connections; thus, adversely affecting the overall accuracy of the chain network method. This conclusion comes from examining the relationship between an actual addressee and his or her position in the reference chain (see Table 1). Specifically, when examining all cases where an addressee is in the reference chain (Columns C and D in Table 1), in 90% of those cases, the addressee is the most recent poster in the reference

Table 1: The Relationship Between an Actual Addressee and His/Her Position in the Reference Chain.

			# of Times an Addressee is NOT in the Reference Chain when Found in . . .		# of Times when an Addressee is IN the Reference Chain as . . .	
			A First Post of a New Thread	A Subsequent Post in a Thread	The Most Recent Poster	Other
Class	# of all Posts*	# of Found Instances of Named Addressees	Column A	Column B	Column C	Column D
Class #1	608	149	50	11	81	7
Class #2	855	271	59	30	153	29
Class #3	1,502	306	37	21	232	16
Class #4	164	96	17	16	51	12
Class #5	412	156	46	26	76	8
Class #6	497	107	27	4	73	3
Average (%)		100%	23%	10%	60%	7%

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

chain (Column C). Thus, it would be correct only 10% or less of the time to consider a person in the reference chain who is not the most recent poster as an addressee of a post.

To determine the exact nature of connections that were missed by the chain network, all posts that correspond to columns A and B in Table 1 were examined for all six classes.

Situation 1: First Post of a Thread

The semi-automated content analysis of posts using ICTA revealed that among the most commonly used names in the first post 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 through some other personal communication such as e-mail. (e.g., *"I just got a reply from [Instructor's name], and she said that [. . .]")*

This type of post is very important in the context of learning, because "student-instructor" ties derived from such messages can be used to identify students who repeatedly ask for the 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 may need extra attention. However, if many students are connected to the instructor through this type of message, it may indicate that lectures or other class materials are unclear to not just one student, and the instructor may need to reconsider either the materials or the delivery method

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 his or her work in the class. This suggests that if there is a tie from an instructor to a student, it is very likely that the student is doing well in the class. Identifying reliable and successful students in a class is an important task, 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 student is doing above average work in the class.

Another common type of message in this category was when an instructor listed groups with their individual members for smaller group discussions. After examining these posts, I concluded that the ties derived from them do not necessarily reflect relationships between the instructor and a student. Instead, these posts 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 participants (See, for example, Haythornthwaite, 2002). Several students confirmed in the comment section of the online survey that they viewed moving 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' posts 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 message is useful in identifying active group members and group leaders when studying collaborative learning. However, group members may perceive too many messages of this type from the same person negatively. For example, in a related study, when analyzing a large collection of Usenet newsgroup messages, Fiore, Tiernan, and Smith (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 message in the online learning environment.

Situation 2: Subsequent Post in a Thread

The detailed examination of this type of message revealed three main types of references or 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 message is likely to connect people who work together. It is also suggestive of stronger personal ties. According to the idea of media multiplexity, stronger ties tend to communicate using more communication channels (See, for example, Haythornthwaite (2001) and Haythornthwaite and Wellman (1998).
- A reference to someone as part of a group when providing 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 message that will likely indicate "work"-related ties.

A reference to somebody who presented or posted something a while ago or using a different communication channel (e.g., *"[. . .] it made me think of the faceted catalogs' display that Susan posted"*). These posts are likely to identify "learning" ties. They show that a poster was not just commenting on the previous post, but rather on something that was said a while ago, thus the poster was following the class discussion, and a student mentioned in the post made a significant contribution to the discussion that resonated with the current poster. All these activities can be categorized as evidence of learning.

In sum, the name network is shown to be well adept at detecting three of the social relations that are considered by many researchers to be crucial in shared knowledge construction and community building: 'help,' 'work,' and 'learning.'

Connections Missed by the Name Network

The previous section summarized common types of connections that were missed by the chain network. However, because on average only 25% of all posts include personal names, it is important to also determine what types of connections were missed by the name network. For this analysis, a small sample was selected from one of the bulletin boards in Class #1. This sample set consisted of 71 posts, of which 43 did not mention any personal names. The manual content analysis of these 43 posts revealed that the majority (31 posts, 72%) did not address any particular person in the class, but rather addressed the class as a whole. Usually, these messages expressed the opinion of the author about a matter under discussion and/or attempted to summarize what had been said in the forum, without reference to any particular person or post. In some cases, a student posted a link that he or she believed to be relevant to the class discussion topic. The results described above suggest that it is possible to ignore messages without names since these messages, for the most part, do not address any particular person in the class. An alternative is to develop a hybrid approach using the chain network and the name network to complement each other. This issue will be the subject of a future study. Further research might also explore how the balance between these more general posts and the name-using posts affect perceptions of how the class functions.

Name Networks and Chain Networks versus Self-Reported Networks

The final part of the study was to compare name and chain networks with self-reported networks to determine which (if any) is a better approximation of self-reported (perceived) social networks. It is presumed that observed social networks, such as chain networks, more accurately reflect relations between group members as compared to self-reported networks, and thus provide a better representation of what is really going on in an online community. But in the context of studying collaborative learning in online learning environments, it may be important to identify and understand perceived social networks. What one student deems important another may value only marginally. Previously, the only reliable way to collect perceived data has been through resource-demanding surveys. An automated method that mimics perceived social networks would, therefore, be a methodological breakthrough.

For this analysis, pair wise comparisons of the three types of networks were conducted using statistical network models and specifically Exponential Random Graph models or just p* models (Robins, in press). To build p* models, XPNET software was used (Wang, Robins, & Pattison, 2006). There are a few important reasons why p* models were selected to conduct this comparison. First, parameters estimated by p* model are easy to interpret and compare across different pairs of networks. Second, the p* model is the only statistical model that is capable of modeling different network structures as well as the 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 chance alone. The results are shown in Table 2. The model was converged (t–statistics < 0.1 for all estimated parameters) and 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 Class #6.

The results show that for four out of the six classes, name networks were consistently more likely to share ties with the self-reported networks than with the chain networks. This supported my general expectation that the name networks would be more reflective of students' perceived relationships. However, for two smaller classes, Class #5 and Class #6, the name networks were less likely to match the self-reported networks. (For Class #6, the model was not significant.) This was a very puzzling but intriguing result. It led to a separate investigation that is currently underway. Early results from this investigation suggest that there

Table 2: EdgeAB—the Likelihood of Two Networks to Share Ties Not by Chance Alone.

	Chain* & Self-Reported Networks		Name* & Self-Reported Networks	
Class	Estimated Parameter EdgeAB	t-Statistics	Estimated Parameter EdgeAB	t-Statistics
Class #1	0.81	0.075	1.73	–0.085
Class #2	0.99	0.044	1.52	0.031
Class #3	1.17	–0.057	1.31	0.001
Class #4	0.61	–0.007	1.11	0.064
Class #5	1.03	–0.004	0.96	–0.071
Class #6	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 Class #4 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.

might be some additional steps that can be taken to further improve the name network method.

First, in addition to connecting a poster with all people who are mentioned in the body of his or her post, the name network method should also connect any two people whose names co-occur in close proximity in the same message. This is because students who work together on the same projects are often mentioned together in the same messages. As a proof of the concept, the name network for Class #5 was rebuilt using co-occurrence of names in the text as an additional indicator of personal ties. This time the likelihood of sharing ties between the name network and self-reported network increased from 0.96 to 1.50 (t-statistics = 0.067), which is higher than the corresponding value from the chain network (1.03).

Second, the 'name network' method should also analyze any other communication media used in a class to better reflect self-reported social networks. This observation is in line with the previous research on media multiplexity, which asserts that people with stronger ties tend to use more types of media to communicate with each other than those with weaker ties (Haythornthwaite, 2001).

Conclusion

For the cases studied, the name network provided on average 40% more information about social ties in a group as compared to the chain network. This additional information is available mostly because the name network can account for instances when a poster addresses or references somebody who has not previously posted to a particular thread.

Furthermore, there is evidence that the name network provides a better reflection of self-reported ties than the chain network. This is primarily because the name network is well adept at detecting three of the social relations that are considered by many researchers to be crucial in shared knowledge construction and community building: 'learning,' 'work,' and 'help.' For example, 'learning' relations were often discovered by the name network method in posts that referred to someone else who had presented or posted something earlier. By referencing ideas or comments from earlier posts, the poster demonstrated that he or she was following the class discussion and learning with and from others. 'Work' relations were commonly revealed by the name network method through posts that referred to an interaction that happened outside the bulletin board, or posts that mention members of the same study group in the same context. Finally, 'help' relations were often discovered in posts from students asking for the instructor's help, or posts that a mention classmate's name in the context of words like "thank you," "help," or "assistance," indicating that a student helped another student. These characteristics of the name network make the method a useful diagnostic tool for educators to evaluate and improve teaching models from the student's perspective. Some specific examples of how ties in the name network can be interpreted and used in the assessment of e-learning are listed below.

(1) To identify students who might need extra attention from the instructor, we can look for students with more connections (abnormally higher tie weight) to the instructor. Or if many different students are connected to the instructor over a short span of time, this may indicate that lectures or other class materials were unclear. This can serve as a signal to the instructor to either cover the materials again, adjust the materials, or change the delivery method.

(2) To find students who are doing well in the class and may be good candidates to provide help to their peers and aid in the learning process of their fellow students, we can look for students who have connections going from the instructor to them. On the surface, it might not be obvious such connections may be used to identify potential student-helpers. However, when you couple the direction of the connection with the knowledge that most instructors tend to reprimand bad students in private and praise good students in public (over the bulletin board), such as complementing a student for some good work in the class, then it becomes clear that such connections are good identifiers of successful students in a class.

(3) To identify students who tend to or would likely to work together on projects, we can look for strong 'student' to 'student' ties. If two students are connected in the name network, it usually means that they are either already working together, or tend to positively view each other's posts and/or class presentations. The instructor can use this information to find students who share similar interests and to group them together for more successful group projects.

(4) To find active group members who often take a leadership role in a group, we can look for nodes that have many connections to other students, especially those connections that were derived from posts mentioning more than one student. These types of nodes tend to be initiated by students who have taken the initiative to organize their fellow students to complete a particular task or objective. The instructor can use this information about potential group leaders to select a contact person in a group or find a person who would be good in leading a class discussion or organizing other class-related activities.

In conclusion, 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, and it is a viable alternative to costly and time-consuming collection of users' data on self-reported networks.

Acknowledgements

The author is grateful to his advisor, Caroline Haythornthwaite, the anonymous reviewers and the JELIS editors, for their help and input during the preparation of this manuscript.

References

Bonnett, C., Wildemuth, B.M., & Sonnenwald, D.H. (2006). Interactivity between proteges and scientists in an electronic mentoring program. *Instructional Science, 34*, 21–61.

Bernard, H.R., Killworth, P., & Sailer, L. (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., & 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.

Fiore, A.T., Tiernan, S.L., & Smith, M.A. (2002). Observed behavior and perceived value of authors in Usenet newsgroups: Bridging the gap. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our world, changing ourselves,* USA, 323–330.

Gruzd, A., & Haythornthwaite, C. (2008a). The analysis of online communities using interactive content-based social networks [Extended Abstract]. *Proceedings of the American Society for Information Science and Technology (ASIS&T) Conference, USA,* 523–527.

Gruzd, A., & Haythornthwaite, C. (2008b). *Automated discovery and analysis of social networks from threaded discussions.* Paper presented at the International Network of Social Network Analysts, St. Pete Beach, FL.

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

Haythornthwaite, C. (2002). Strong, weak, and latent ties and the impact of new media. *The Information Society, 18*(5), 385–401.

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 & S. Oksefjell (Eds.), *Out of corpora: Studies in honour of Stig Johansson.* Amsterdam: Rodopi.

Matsuo, Y., Hamasaki, M., Nakamura, Y., Nishimura, T., Hasida, K., Takeda, H., Mori, J., Bollegala, D., & Ishizuka, M. (2006). Spinning multiple social networks for semantic web. *Proceedings of the Twenty-First AAAI Conference on Artificial Intelligence, USA,* 1281–1387.

Nurmela, K., Lehtinen, E., & Palonen, T. (1999). Evaluating CSCL log files by social network analysis. *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning, USA,* 434–444.

Reyes, P., & Tchounikine, P. (2005). Mining learning groups' activities in forum-type tools. *Proceedings of the 2005 Conference on Computer Support for Collaborative Learning: Learning 2005: the Next 10 years!, Taipei, Taiwan,* 509–513.

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

Snijders, T.A.B. (2008). *Models for social networks.* Retrieved January 10, 2009 from University of Oxford, Department of Statistics Web site: http://www.stats.ox.ac.uk/~snijders/sna_course.htm

Wang, P., Robins, G., & Pattison, P. (2006). *PNet: Program for the estimation and simulation of p* Exponential Random Graph Models,* User Manual. Retrieved January 10, 2009 from University of Melbourne, Department of Psychology Web site: http://www.sna.unimelb.edu.au/pnet/download/PNet/PNetManual.pdf

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), 46–56.