Analyzing Networked Learning Texts

Caroline Haythornthwaite, Anatoliy Gruzd

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

Abstract

Social interactions are essential in understanding the collaborative processes in networked learning environments. Although individuals may learn by retrieving information from online archives, dictionaries and encyclopaedia, it is the interaction with others with similar, perhaps narrowly enjoyed interests that fuels the benefits of networked learning. This paper presents our ongoing work on a novel, automated method for extracting interaction data from threaded discussions of networked learning groups. Using natural language processing, the proposed method reduces large text-based datasets to community and conversational essentials that show the relations of importance to group members. By studying these relations, we hope to identify what matters in terms of learning in the online interaction space and to provide useful representations of online conversations to help networked learners (instructors and students) better understand the social environment in which they are participants. To do so also requires making accurate determinations of who is talking to whom. This paper discusses the methodological issues associated with extracting names from networked learning texts and our procedures for enhancing network information through new techniques of name extraction.

Keywords

Social networks, natural language processing, collaborative learning

Introduction

The term 'network' has many meanings. It can be applied to describe computing hardware infrastructures and the connectivity between computers that these provide, in local area, internet, and wireless networks that provide in-house to global communication on desktop, laptop and mobile devices. It can refer to software programs that provide a platform on which individuals can form interpersonal ties, such as listservs, chat rooms, online learning environments, and social software/social networking services. And, it can refer to the networks of colleagues, friends and family that make up our social worlds. What unifies these applications of the word 'network' is the common idea of individual nodes (computers, individuals, organizations) tied by some property that forms a structure greater than the sum of its parts. In social network analysis terms, the nodes are *actors* connected by *relations* that form *networks* that reveal the patterns of interconnection that sustain the whole *social network*. It is social because it is interactions between actors that create the network, rather than some physical or technological connection. Interaction provides the essential substrate of social networks; they do not exist without some social actor taking part in the relations that connect them to others.

Interaction is also essential in many approaches to networked learning. Although individuals may learn by retrieving information from online archives, dictionaries and encyclopaedia, it is the possibilities of interaction with others from around the globe with similar, perhaps narrowly enjoyed, interests that fuels the benefits of networked learning. A social network view is in keeping with notions of collaborative learning, participatory culture, web 2.0, and learning through engagement with others (Bruffee, 1993; Cook & Brown, 1999; Jenkins, 2006; Koschmann, 1996; Lave & Wenger, 1990; Miyake, 2007). Thus, examining social networks – including the roles and positions of actors in a social network, their influence on others, and what exchanges support and sustain the network – is an important goal for understanding networked learning processes.

However, pursuing a social network approach raises a number of methodological issues. The first is how to

ISBN No: 978-1-86220-206-1

examine and evaluate the network aspects of networked learning, including identifying what matters in terms of learning in the online interaction space. The second issue is how to do this on a scale that is adequate to give more than anecdotal results, and which keeps pace with the rapid production of text typical of networked learning settings. Recent estimates of online text production report massive amounts of data: 610 billion to 1100 billion email messages in 2000 alone, with an average size of 18,500 bytes, with "the amount of flow becomes surprisingly gigantic, somewhere between 11,285 and 20,350 terabytes" (http://www2.sims.berkeley.edu/research/projects/how-much-info/internet.html). Technorati currently tracks 27.2 million weblogs, with approximately 2.7 million users updating their blogs per week, and the size of the blogosphere doubling every 5.5 months (http://www.sifry.com/alerts/archives/000419.html). The numbers are also there in students and courses. In the U.S. alone, over 3.5 million students were studying online in the Fall of 2006, and 20% of U.S. higher education students took at least one course in Fall 2006 (Allen & Seaman, 2007). Each of these venues, plus chat and listserv traffic, produce an information and network trail relevant to networked learning.

While online interaction is creating this growing legacy of texts, there are few mechanisms available for processing and analyzing such data and connecting these to performance, learning or social outcomes. Patterns of posting, interactivity, and topic threads hold information pertaining to group identity, growth and maintenance; use of words, short-hands and acronyms show the extent to which groups are embedded in disciplinary or professional norms of discourse; and interaction patterns, topics, and style provide important pointers to group purpose and conduct. (For work in this area see, Carley, 1997; Carley & Palmquist, 1992; Corman, Kuhn, McPhee & Dooley, 2002; Gloor & Zhao, 2006; Haythornthwaite & Gruzd, 2007; Rafaeli & Ariel, 2007; Turner, Smith, Fisher & Welser, 2005; Webster, 2007; Welser, Gleave, Fisher & Smith, 2007).

To draw interaction data from these texts, two steps are needed. First, some form of automated processing is needed to reduce the large datasets to community and conversational essentials that show the relations of importance to group members; and second, assessment of these data extractions is needed to determine the usefulness and meaning of these measures to participants. When combined, these two aspects can provide useful representations of online conversations, from statistical reports to visualizations of data and interactions, each of which can help networked learners (instructors and students) better understand the social environment in which they are participants.

This paper reports on our work addressing these two components. A technological infrastructure known as the Internet Community Text Analyzer or ICTA for short (pronounced 'ishta'), developed by Anatoliy Gruzd, provides an environment for natural language processing of online texts and extraction of interaction patterns. Natural language processing is used to extract actors and topics from the online conversations, and conversational records are analyzed to extract posting behaviour, conversational interaction, and network formation.

Although the tool can address any online dataset, we have focused on class-wide discussion boards from eight interactions of a single online course, taught by the same instructor. We are examining this dataset in detail to test the practicality and usefulness of the results for instructors and researchers interested in understanding online behaviours. Bulletin boards are used widely in online classes, internet discussion groups, and online communities. Thus, we believe that determining ways of analyzing this particular form of discussion will have wide applicability for learning contexts, including those independent of traditional educational settings.

In our presentation we will demonstrate the ICTA tool (which exists as a working prototype), discuss findings, and methodological issues associated with this kind of analysis. What follows reports on the processes devised for identifying actors in the dataset. Results relating to the online learning classes represented in our dataset will be presented during the presentation.

Dataset

The dataset includes all class-wide postings from bulletin boards for a required class for first term library and information science students taking their Masters degree online. Classes typically meet weekly in 'live' online sessions conducted through a combination of internet supported video, slides, and chat. Bulletin boards are then used over the week for discussion of topics initiated by the instructor. The class-wide bulletin boards are password protected and thus not openly available on the web. Postings from eight iterations of the class are available, each given by the same instructor, two per Fall 15-week term from 2001 to 2004. The eight classes involved 31 to 54 class members, the professor, and 3-4 teaching assistants. Together they posted 1200 to 2100 class-wide messages per term. Students also had small-group bulletin boards in use during these terms, and posted 2-3000 messages a term on these boards, but for privacy reasons these are not part of the dataset. The online learning system used at the time was an in-house application created and supported by the degree-granting school. Beyond bulletin boards, the community was also maintained via other online means, including email, and online chat during live weekly class sessions, and one campus visit per term. Table 1 gives the basic statistics for the four courses. Participants – students, instructor and 4-5 teaching assistants – posted 1207 to 2157 messages on the class-wide bulletin boards during the 15-week period.

Institutional Review Board permission was obtained for this work; procedures included alerting the class to the intended use of the data in the class-wide bulletin boards and describing the intended use. Students were given the option of contacting the researcher directly if they did not want to be directly quoted from the bulletin boards.

Table 1: Basic statistics for class-wide bulletin board postings, eight classes

	F01A	**F01B**	**F02A**	**F02B**	**F03A**	**F03B**	**F04A**	**F04B**
No. of Messages	1207	1581	1469	1895	1279	1242	1493	2157
No. of Participants	38	47	47	54	54	46	54	52
No. of Bboards	22	22	28	28	25	24	28	27
Avg.No.Symbols/Msg	1073	1056	864	898	1286	953	967	1058
Avg. No. Lines/ Msg	17	14	15	14	17	17	15	16

Determining Networks

In analyzing networked learning environments our aim is to make visible the interaction dynamics that are hidden in streams of linear text. Since interaction requires identifying participants, the first problem to solve is how to identify the actors, and then to derive "who is talking to whom." Later we want to add to that "what" they are talking about.

Typically, all that is evident as an overview of the list of bulletin board postings is the email address (or other identifier) of the poster, and the subject line. At first glance this seems to provide a simple mechanism for identifying who is talking to whom – a poster is answering the previous poster. This is one way to build the network, and we refer to this network as the *Chain Network* – built on the way messages chain to each other temporally.

Building Chain Networks

In constructing networks from header information, there are still some decisions about the relevance and weight of earlier postings that need to be made. Table 2 presents some options. The overall question is what measure of influence or prominence should previous postings be given in considering the tie between posters. The basis of the social network perspective is consideration of the way each individual's behaviours affect the thinking and behaviours of others. We can readily expect that a first poster will influence postings that follow because of his or her primacy in addressing a topic presented for discussion. This post gets the ball rolling, provides an opinion to respond to, and discussion norms (or instructions) may dictate that subsequent posters pay attention to earlier postings. In promoting participation in online discussion, instructors may actively discourage non-reflective postings, i.e., discussion boards are not just sites to submit individual assignments, but are instead places of activity, of to and fro of discussion. Although not all boards in all settings will be used as, or be successful as interactive, participatory discussion spaces, one thing every instructor would want to know if whether they have been used in this way. The chain data alone cannot actually determine this, and we return to ways of interrogating the actual interactive process below. Suffice it to say for now that in building the most representative chain network in what is expected to be a reflective, discussion-based forum, requires some consideration of the relation among previous posts.

In our formulation, we have considered several options. *Option 1 is the naïve solution, creating the network based on counting a tie to be present only between a message poster and the poster of the immediately preceding post.* In the options discussed here, the ties are treated as ***undirected***, i.e., discarding the direction of the connection between actors. In accepting undirected ties, we have reasoned as follows. We assume that an individual who first posts does not set out to influence any particular individual member of the class. Thus, the receiving node for a posting directed "out" from a first poster cannot be determined. Although such information might be in the post itself, for the chain network, using header information only, we do not want to assume more than is indicated by the available data, and thus we assume only a general intention to influence or engage with the class as a whole.

A responder does answer an individual, i.e., the previous poster, but we do not know without analyzing the text whether that response is an acknowledgement of influence, a challenge to it, or a completely new point, unaddressed to the previous poster. Thus a determination of an intentional directed tie from the responder to the previous poster cannot be made unequivocally from position in the posting sequence. Thus, we reason instead that juxtaposition is a sufficient, instrumental, indicator of a tie, but direction of the tie cannot reasonably be assumed. Thus a tie "in" to the previous poster is not assumed, only a tie based on sequencing within the message stream. Neither can a tie "out" from a secondary poster to subsequent posters be made for the same reasons as no tie "out" from the first poster is determined. Like the first poster, subsequent posters' influence extends to the class as a whole, and to subsequent, yet-to-be-heard posters as well.

Options 2 and 3 bring in consideration of the chaining effect of streams of posts. *Option 2 considers only the influence of the first poster as the prime mover of the discussion.* A weight of constant value, equal to or less than that assigned to the immediately preceding poster is assigned to the tie. Option 3 considers the influence of all previous posters, using weights of ordered diminishing size to weight the tie between the poster and all previous posters

Table 2: Chain network options

Options	Amy ← Bob ← Cathy ← David		
(**1**) Connect a sender to the immediately preceding poster only (undirected), e.g., a connection is counted only between poster David and Cathy is counted	0	0	1
(**2**) Connect a sender to the first (=thread starter) and immediately preceding poster, assigning a weight to both ties, e.g., a connection is counted between poster David & Amy, and David & Cathy.	<=1	0	1
(**3**) Connect a sender to *all* posters in the reference chain, assigning weights that decrease with 'distance' from the poster (e.g., reducing each by half)	.25	.5	1

Shortcomings of Chain Networks

These options are our starting point for examining interaction networks. They represent a logical set of criteria for building networks based on the posting chain information only. However, while these procedures provide some approximation of the conversational progress, there are a number of shortcomings of these techniques. In an asynchronous, many-to-many discussion board, individuals may address messages that appear much earlier than the immediately preceding posting, making the chain data a poor estimate of network interaction even if an accurate representation of the chain left in the textual artefact that results from use of the discussion boards. An individual may post in one apparent place in the message sequence, but refer in their message to one of more of the postings preceding their post, or to conversations and discussions that happened outside the discussion board (e.g., in our case, during the live sessions, or in different bulletin boards on the class discussion board). Further, an individual may seem to respond to one post, but in their text refer to several others, synthesizing and bringing together comments of others.

Examples from our dataset are given in Table 3 (names have been anonymized). The first example shows four individuals named in the text of the message who are directly addressed by the poster (Nick, Ann, Gina, Gabriel), but using only the previous poster information only one name (Gabriel) would be included in the network. The second example shows an unambiguous addressee (Gina) but if the network is built from ties across the entire chain history, extra people would be included in the network (Gabriel, Sam, and Eve as well as Gina); and if Gina is not the immediately preceding poster, a connection might not even be made to her. The third example names a person who has not posted at all in this thread, and hence would not be identified at all by a chain network.

Table 3: Examples of differences between chain data and text data

Chain	**Text**
Previous post is by Gabriel, Sam replies:	'**Nick, Ann, Gina, Gabriel**: I apologize for not backing this up with a good source, but I know from reading about this topic that libraries…'
Previous posts by Gabriel, Sam, Gina, and Eva, then:	**'Gina**, I owe you a cookie. This is exactly what I wanted to know. I was already planning on taking 302 next semester, and now I have something to look forward to
Post by Fred:	'I wonder if that could be why other libraries around the world have resisted changing – it's too much work, and as **Dan** pointed out, too expensive.'

Each of these kinds of shortcomings in the chain network leads us to look at the text of the message for more detail on who is talking to whom, and about what. Our second approach uses natural language processing to identify and extract names from the text of the messages in order to build the who-to-whom network.

Name Networks

Identifying individuals from names within the text is not a straightforward issue for automation. Although we may have a master list of enrolled students, differences between names used and class lists is a common issue: Virginia becomes Gina; Michael John Smith or Michael J. Smith goes by John; Wendy Mason became Wendy Carpenter last term but here record remains in her former name; the instructor is identified as "Professor" rather than by their name; a student acquires and is referred to by a nickname ("JJ", MaryK) students with the same first name start being identified separately at some point in the term (e.g., Mary Kipley and Mary Donnelly both appear as Mary early on, becoming Mary K and Mary D later posts).

To explore the variations that might be present, we hand coded a bulletin board containing 62 messages. This revealed a Number of issues and conventions about the use of names. First, four categories of name use could be distinguished: those referring to participants in the class; those to non-participants, most commonly the author of a work under discussion; names appearing because of errors, e.g., incorrectly spelled names; and names appearing in the copy of an earlier posting appended to the new post. Table 4 gives some further specifics of names uses found within these categories.

Table 4: Name occurrences in bulletin board discussions

Usage	**Network Participants**
From	Person indicated in 'from' line of heading, always an email address (system generated)
Addressee	Direct reference to other ('I agree with you Todd')
Reference	Indirect reference to other ('Todd has a good point')
Self-Reference	Poster refers to themselves in some way (brain-dead student, high school teacher)
Signature	Name as given by the message author on their post
	Named Non-Participants
Subject	Subject of the discussion, 1-3 parts, e.g. Dewey, Brewster Kahle, Charles R. Darwin
Non-Group	Person not in the group, nor the subject, e.g., a former professor or mentor
	Errors
Error	New name appears because of error (e.g., Lackie as a subject instead of Leckie)
	Previous Posts and Copies
[Previous Posts]	If the previous message is included, indicates the previous poster ('Janice wrote: …') (system generated; could be edited deliberately or accidentally when)
[Copy]	Name appears because it is included within the previous message

Note: For our dataset, postings are first processed to remove copies of earlier messages. However, for some analyses copying behaviour may be an aspect for investigation and hence names from copies would matter.

What needs to be created in these cases is an authority file that ties name variations, nicknames, and incorrectly spelled names to a single identifier (e.g., email address). From this and further examination of the bulletin boards these steps were identified to be dealt with for automated detection of network participants from names used in the dataset:

1. Disambiguating names and nicknames from other text
2. Disambiguating names of people from names of people being discussed (e.g., subjects of discussion)
3. Detection of aliases for a given person (multiple names for a single person)
4. Disambiguation of two or more users with the same name (multiple people for a single name)

Automated Node and Tie Discovery

The method devised to automatically identify node and ties uses text analysis and assigns probabilities to discovered relationships in accordance with the likelihood that what has been identified is a name. In short the method entails (1) extracting names from the dataset and assign a probability value to each according to the likelihood it is a name; (2) determining a relationship between an email address and the name; and (3) determining the ties among email addresses and assigns a weight to each discovered tie.

In step 1, probabilities given to extracted names are higher when the name is (1) capitalized, and (2) found in the index of personal names as found in the listing for the 1990 US Census (http://www.census.gov/genealogy/names). Both of these choices are specifically tied to finding names of US participants. There are some shortcomings in these even for US classes given the mix of students from many countries. To overcome this inherent weakness, census data from other country or countries can also be incorporated into the program. However, too many name lists will likely increase false positives. To recognize names that are not yet in the dictionary (e.g., nicknames, abbreviated names, unconventional names, etc) such as *CH* or *CarolineH*, we relied on the context words that usually precede personal names such as titles (e.g., *Professor*, *Major*, *Ms.*) and greetings (e.g., *Hi* or *Dear*). Quantification of false positive and negative name identifications from this technique needs to be included in future work. (For work on automated personal name discovery, see Harada, Sato & Kazama, 2004; Minkov, Wang & Cohen, 2005; Patman & Thompson, 2003; Sweeney, 2004.)

In step 2, names are associated with email addresses. The essence of this process is to identify whether a name is more likely to belong to the poster of the message or an addressee mentioned in the text of the message. This is accomplished by relying on *structure-based* (e.g. word position) and *content-based* (e.g. context words) information about the names found within the postings. For example, names appearing at the beginning of a post are more likely to be those to whom the message is addressed; names at the end, more likely to be signatures and thus from whom the message is sent. Additionally, if a name appears in a close proximity (within 1-2 words) to one or more words or phrases that commonly appear with addressees, we increase the confidence level of that particular name-email association. Examples include greetings (e.g., “hi”, “hello”, “dear”), agreement (e.g., “agree/disagree with”), referencing others (e.g., “according to”, “said that”). We apply the same confidence level increases to phrases that are commonly used in the signature such as ‘thank you’, ‘best regards’, ‘cheers’. Figure 1 provides an example (see Gruzd & Haythornthwaite, 2008 for a more detailed description of this algorithm.)

In step 3, the association between individuals is made to create the ties for the network. First, each sender in the class is tied to all names mentioned in his/her emails. For example, the name ‘Wilma’ from Figure 1 is associated to a unique email when she posts a message. Within that message, Wilma is tied to Dustin by naming him, and that name has been found to be associated with the email tank123@gl.edu. Wilma is also tied to Charlie, but as there is no email associated with “Charlie” this is taken to indicate a person who is not part of the current class, e.g., ‘when Charlie and I took Professor Sid’s course last year’. Lastly, Wilma may appear with no name associated. In this case, the only information on ties is present in the chain network. This may occur when Wilma is the first poster to a thread, or if she makes a post without naming anyone else.

Figure 1: Example of name identification and association with email address

From: wilma@bedrock.us (=**Wilma**)

Reference Chain: tank123@gl.edu (=**Dustin**) => hle@gl.edu (=**Sam**)

Hi Dustin, **Sam** and all, I appreciate your posts from this and last week […]. I keep thinking of poor **Charlie** who only wanted information on "dogs" […] Cheers, **Wilma**.

Words to the Left	Name	Words to the Right	Position %	Context Word	Is Poster?	Is Addressee?
Hi	**Dustin**	Sam Nick	0	Hi	0.00	**2.00**
Hi Dustin	**Sam**	Nick and	1	Hi	0.01	**1.98**
of poor	**Charlie**	who only	50		0.50	0.50
Cheers *	**Wilma**	*	88	Cheers, *	**3.52**	0.12

* symbolizes a new line

Note: If the difference between two values in the last two columns of the table above is insignificant (less than 10%), then we ignore the instance of this name completely due to a lack of information to determine whether a name is a poster or an addressee in the message.

Armed with the chain and name network, it is then possible to view the resulting networks. Figures 2 and 3 show a subset of the name and chain network for postings in three bulletin boards, the egocentric network for the actor Brent. Overall the chain network for the sample set found 37 nodes in 346 messages. The chain network has 223 ties, and the name network 215 ties of which 140 are the same (r=.453). Thus, although there is considerable overlap, there are also substantial differences in what is revealed by these two network derivations.

To better understand these differences, we compared all connections that make up each tie from the name network with those of the chain network. For this task we selected a larger sample set of 534 messages. Of these messages, 315 messages were of a general nature and did not indicate strong social ties between group members. However, the remaining 219 messages contained explicit references to other people in the group. In analyzing these 219 messages, our program found 280 explicitly identified names of addressees that were then used to build the name network. We discovered that 108 (38.57%) of all addressees were not in the reference chain. This means that for these 219 messages, regardless of the method used for building it, a chain network misses about 38% of potentially important connections and about the same amount will be incorrectly identified.

Egocentric visualizations that include as nodes named entities not associated with emails reveal aspects of name detection that may need to be addressed with refinements to the name detection algorithms. In particular, names of places and institutions need to be considered as these are appearing as names referred to by individuals, even if not associated with emails. Similar issues are occurring in distinguishing names from concepts, e.g., does "mark" indicate a person or a process ("mark-up"). Although these may be seen as 'errors' in the detection algorithm, they also show how the social network approach can make evident the name detection process and help in its refinement. It is likely that no algorithm will perfectly detect nodes. Thus, strategies for narrow or wide casting of nodes in necessary to favour excluding or including false positives. We expect that an individual may choose a wide case, and then hand correct for the final name list. The ICTA software tool currently allows for this kind of refinement toward the name list. But it is important to note that the hand coding noted above took 3 hours to complete, and the ICTA algorithms take about 3 minutes. Thus, hand correcting is a much less onerous task than coding by hand.

Summary

These chain and name network approaches described here begin the task of understanding and extracting social interaction networks from discussion board data. However, they are only the starting point. Our future plans call for comparing the chain and name networks to each other, and to networks generated from participant judgements of interactions. Further, we foresee refinements to both the chain and name methods of defining networks given feedback from participants at the network conferences, as well as from instructors and students who take classes online and engage in online discussions.

ISBN No: 978-1-86220-206-1

References

Allen, I.E. & Seaman, J. (2007). *Online nation: Five years of growth in online learning.* Needham, MA: Sloan Consortium.

Bruffee, K. A. (1993). *Collaborative learning: Higher education, interdependence, and the authority of knowledge*. Baltimore: John Hopkins University Press.

Carley, K. & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social Forces, 70*(3), 601-636.

Carley, K.M. (1997). Extracting team mental models through textual analysis. *J of Org. Beh.,18*(S1), 533-558.

Cook, S. D. N, & Brown, J. S. (1999). Bridging epistemologies: The generative dance between organizational knowledge and organizational knowing. *Organization Science, 10*(4), 381-400.

Corman, S., Kuhn, T., McPhee, R., & Dooley, K. (2002). Studying complex discursive systems: Centering resonance analysis of communication. *Human Communication Research, 28*(2), 157-206.

Gloor, P.A. and Yan, Z. (2006). Analyzing actors and their discussion topics by semantic social network analysis. *Proceedings of 10th International Conference on Information Visualization*, London, England.

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

Harada, M., Sato, S. & Kazama, K. (2004). Finding authoritative people from the web. *Proc. Dig. Libraries,*.

Haythornthwaite, C. & Gruzd, A. (2007). A noun phrase analysis tool for mining online community. In C. Steinfield, B.T. Pentland, M. Ackerman & N. Contractor (Eds.), *Communities and Technologies 2007* (pp. 67-86). London: Springer.

Jenkins, H., with Clinton, K., Purushotma, R. Robinson, A. J., & Weigel, M. (2006). *Confronting the Challenges of Participatory Culture.* Chicago, IL: MacArthur Foundation.

Koschmann, T. (Ed.)(1996) *CSCL: Theory and practice of an emerging paradigm*. Mahwah, NJ: Erlbaum.

Lave, J. & Wenger, E. (1991). *Situated learning*. Cambridge, UK: Cambridge University Press.

Minkov, E., Wang, R.C. & Cohen, W.W. (2005). Extracting personal names from email: Applying named entity recognition to informal text. *Proceedings of Human Language Technology Conf.* Vancouver, BC.

Miyake, N. (2007). Computer supported collaborative learning. In R. Andrews & C. Haythornthwaite, *Handbook of E-learning Research* (pp. 263-280). London: Sage.

Patman, F. and Thompson, P. (2003). Names: A new frontier in text mining. *Intelligence and Security Informatics*. First NSF/NIJ Symposium, Tucson, AZ.

Rafaeli, S. & Ariel, Y. (2007). Assessing interactivity in computer-mediated research. In A.N. Joinson, K.Y. A. McKenna, T. Postmes & U.D. Rieps (Eds.), *Oxford Handbook of Internet Psychology* (pp. 71-88). Oxford, UK: Oxford University Press.

Sweeney, L. (2004). Finding lists of people on the web. *ACM Computers and Society*, *34*(1).

Turner, T. C., Smith, M. A., Fisher, D., & Welser, H. T. (2005). Picturing usenet: Mapping computer-mediated collective action. *JCMC, 10*(4), article 7. http://jcmc.indiana.edu/vol10/issue4/turner.html

Webster, A. (2007). *Visible relations in online communities: Modeling and using social networks*. Master's Thesis. University of Saskatchewan. http://library2.usask.ca/theses/available/etd-09192007-204935/

Welser, H.T., Gleave, E., Fisher, D. & Smith, M. (2007) Visualizing the signatures of social roles in online discussion groups. *JOSS, 8*(2). http://www.cmu.edu/joss/content/articles/volume8/Welser/.

Figure 2: Ego network for Brent: Name Network

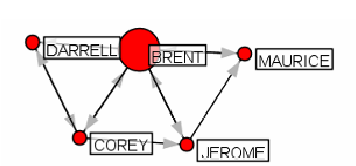
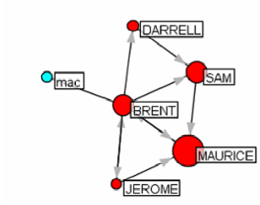


Figure 3: Brent: Chain Network



ISBN No: 978-1-86220-206-1