Network Analysis of Trace Data for the Support of Group Work: Activity Patterns in a Completely Online Course

Sean P. Goggins
Drexel University
Philadelphia, PA 19104
outdoors@acm.org

Krista Galyen
University of Missouri - Columbia
Columbia, MO
kgalyen@gmail.com

James Laffey
University of Missouri - Columbia
Columbia, MO
laffeyj@missouri.edu

ABSTRACT
A16-student, completely online software design coursewas studied using social network analysis and grounded theory techniques.Bi-directional (read and post) log data of user activity was recorded to understand how small group networks change over time with activity type (individual, peer-to-peer, and small group). Network structure was revealed through sociograms and triangulated with discussion board topics and interview data on group experience. Results show significant differences in network structure across activity types, which are supported by open coding and axial coding of the text of member discussions and editing patterns of member work products.It is also indicated that bi-directional log data, contextualized to specific activities and artifacts, revealed a more accurate and complete description of small group activity than ordinary, uni-directional log data would have.Our findings have implications for tool development revealing group structure and software design for completely online group work.

Categories and Subject Descriptors
H5.3. [Group and Organization Interfaces] Computer supported cooperative work.

General Terms
Measurement, Documentation, Performance, Design, Human Factors, Theory

Keywords
SNA, CSCW, CSCL, Group Development, social capital, human capital, networks of practice, groups, teams, communities.

1. INTRODUCTION
Small groups are the "engines of knowledge building"[21], and their ability to innovate and construct knowledge has been widely acknowledged.Completely online learning is now commonplace in higher education[1], and gaining popularity across the spectrum of human interaction [8]. Small groups who come together completely online are an important and emerging form of social interaction, but remain little examined in the Groupliterature.

Prior studies that examine completely online small groups do so with limited tools. For example, social network analysis (SNA) is a powerful tool for analyzing the social interactions of small groups online. However, uni-directional log data is currently the dominant form of capturing and analyzing these interactions [2,14]. In this study, we address these two interwoven problems: 1) how different activity types shape the social organization and practices in *completely online*computer supported collaborative learning (CSCL) groups, and 2) how *bi-directional log data*(e.g.user posts *and* reads) – as opposed to uni-directional data (e.g. user posts only) – makes the activity types and group interactions more completely visible. (See section 3.3.1 for a full explanation of the bi-directional log data used in this study.)

GROUP'10, November 7–10, 2010, Sanibel Island, Florida, USA.

1.1 Making Practice Visible Online
The differences between completely online and face-to-face interactions frame the experience of the groups and group members we study.Completely online groups experience design features as opportunities to work and learn in new ways, but also as potential social barriers that may hinder successful collaboration and in turn negatively impact knowledge building.Studying the practices of completely online groups as they interact and develop can help identify factors or aspects of performance in these contexts that lead to successful interactions and prevent failure, a requirement for successful computer-supported collaborative learning [20].

However, to accurately understand the practices of these completely online groups, a set of analysis techniques should be applied which reflect learner activities as accurately and completely as possible. An increasingly common form of analysis in online settings is social network analysis, where the online group logs are analyzed to reveal the structure of relationships.Howison, Wiggins&Crowston[11] point out that online group logs provide a richer, more complete record of the traces of online interaction than was available to sociologists who first sought insight through structural analysis. However, SNA theories built upon decades of research in the physical world are not consistently applied in online settings[11]. Finding the type of data that most accurately represents online interactions is needed.

To date, the dominant form of online social network analysis data comes in the form of uni-directional data (such as posting in a thread) rather than bi-directional data (such as both posting and reading a thread) [2,14]. While the significance of "read data" (also known as "passive" activity) hasnot been generally recognized as a vital component in online social network analysis, the act of observing others as a part of learning and interacting is widely acknowledged as integral and important [15]. Thus in order to accurately represent both the "passive" and "active" components of learner activity, both data types ought tobe includedin activity data logs.

The inclusion of more representative data could help build a more complete picture of what is happening in the online environment. Students and instructors are often not fully aware of the passive activity of reading in an online course, and broader patterns of activityremain invisible. Present tools do not make it easy to observe what is going on in a completely online course.However, bi-directional log data can reveal this activity and is able to reflect a learner's activity more realistically.For example, it would be more accurate to say, "John has read these 20 of his classmates' work products and responded 3 times to 2 different mates" than "John responded to two classmates." Having more accurate and richer data will help build understanding of completely online small groups. Analysis of this data will help to lift what has been calledthe "black veil"over activity in online learning.

In this study we analyze a single completely online course of 16 people and show how differences in activity type correspond with differences in network structure. The use of bi-directional SNA log data is integral to accurately understanding and portraying the network structure in these groups.

2. SNA APPLIED TO CSCL TRACES

2.1 Group Structures Over Time

In the CSCL literature, the shift from taking a snapshot of present group state toward understanding how online groups change over time is addressed from two methodological perspectives. First, by de Laat et al's[14] use of social network analysis as a research method and second, by Cress et al's[5] introduction of hierarchical linear modeling (HLM).De Laat et al applied SNA to a context-poor set of CSCL trace data, demonstrating the basic utility of SNA methods for CSCL log analysis.Cress noted in her study of the utility of HLM for studying collaborative groups that while there are indeed differences within and between collaborative groups, larger differences (or variance) can be found in the different time periods or themes in which the groups reside.This suggests that online group performance may be influenced by activity and task, and multiple snapshots should be taken over time to reflect how context evolves with group structure.

In addition, facilitating successful member interactions as activities and the structure of groups change is critical to online learning.Therefore increasing the visibility of social interactions within and across groups calls for analysis of structural change over time.One way to accomplish this is by understanding how social structure varies by activity type.To facilitate these successful interactions, instructors and members require a view of how they are working togetherand how different SNA structures and measures relate to group performance and cohesion over time.Little research is presently focused on completely online CSCL groups, online small-group structures or online, small-group activity patterns.In addition, no prior research has examined group structures and contrasted those structures across activity types in a completely online context.

These gaps in the research lead us to several questions. We would expect the structure of online groups to vary by activity type, but do they?If so, how does this variation come into view?What role does activity type play in this variance?And is there a relationship between patterns observed in logs and the experiences described by participants?Working to answer these questions will illuminate the nature of completely online small group collaboration, the factors that facilitate successful collaboration and the different social network patterns that emerge from different activity types.

For most previous social networking research in environments outside CSCL, SNA has been used to analyze interactions through a single post-hoc visualization rather than looking at interaction patterns over time [12].Temporal network analysis helps to build understanding of how context is constructed through member participation.Katz et al [12]theorized that SNA snapshots at numerous points in time during collaboration would shed light on the nature of group development and context.

To better understand online CSCL groups, Katz et al's idea of time series network analysis needs to be applied. Looking at small groups over different times periods and annotating the different activity typeswould be a step forward inunderstanding completely online group development.Analysis of fine-grained bi-directional log data, combined with qualitative analysis of member generated discussion and content can provide a rich understanding of online group developmentacross different activity types.

2.2 SNA: MakingOnline Structure Visible

2.2.1 A Review of SNA Studies

All the studies mentioned in the following review use social network analysis to understand online interaction patterns. In addition to summarizing the contributions they make to framing our study, we highlight three important aspects: 1) the use of uni-directional or bi-directional log data in social network analysis and subsequent visualizations of structure, 2) the number of data "snapshots" used to see changes in group structure over time, and 3) the unit of analysis (communities versus groups).

Aviv, Erlich&Ravid[2] useduni-directional log data from a single point in time to compare the structure of two different communities.They performed social network analysis of two CSCL communities by looking at the cohesion and roles within those communities.They found that knowledge construction by course members was higher in structured online groups than it was in unstructured online groups.They also found evidence of identifiable interaction patterns through SNA.For instance, dense networks produced more knowledge.However, the highly controlled, experimental nature of their study limited generalizations to naturalistic environments.Computer mediation in a lab is distinct from computer mediation in a geographically dispersed, completely online lived experience.Due to the single-point-in-time and community-oriented nature of the study, changes in the structure within communities over time were not reported.Haythornthwaite[10]highlighted the potential of viewing patterns of interaction as determined by activity type yet acknowledged the confounding effects of team structure.These two studies, together, highlight a gap in the literature related to our understanding of changes in group structure during different activity types. Closing this gap willadd clarity to understanding groups in completely online CSCLsettings.

De Laat et al [14] established the utility of SNA in CSCL studies as a means for triangulating the findings derived from other methods and adding interpretive richness.For example, they showed that instructor position in online social networks was visible through SNA, and that very active and nominally active members of the course were easily distinguished.De Laat et al [14] useduni-directional log data from three points in time to study interactions within groups in a networked learning community.They captured posts and responses in a Web-CT-hosted, masters-of-education learning community and integrated their analysis of the network with content analysis and critical incident interview data.Their study used member post data to create sociograms of the social network in the course, and measured the centrality and density of the social network.Post data described active response by members to each other, but did not include a record of passive activity like reading the work of others.This uni-directional nature of the logs used by de Laat et al [14]provided a narrow view of one type of interaction, and the log record was not contextualized to specific topics, discussion boards or posts in the way the data are in the current study.For example, network analysis of uni-directional logs waslimited because a participant who posts 70 items and reads 3 items shows the same level of social connectedness as a participant who makes 70 posts but reads 350 posts.The activity that is invisible in uni-directional logging is key to our understanding of how the structures of CSCL groups unfold.

Shen et al [19] used bi-directional logs of read and post activity at one point in time to describe how interaction influences sense of community in an online learning environment.They compared two completely online courses focused on the design of collaborative software systems, and taught by the same instructor.Shen et al [19] showed that the development of a sense of community in each course is different, and that these differences in sense of community corresponded with differences in social network structure.The current study extends beyond Shen's work by capturing(1) the network at different points in time, (2) the antecedent poster identities in each discussion board, and (3) the identity of the original discussion board creator.

2.2.2 Making Online Structure Visible

Analysis of how groups of people relate to each other and subsequently evolve into groupings of different sizes is a type of structural analysis in the social sciences.Structural analysis makes network patterns visible, statistically and visually, in the form ofsociograms. Wasserman & Faust[25] and Koku& Wellman [13] describe social network analysismethods for understanding these structures.The core concepts represented by any SNA method are the actor (who can be an individual, group or event), the relational tie (between actors), dyads (two people), triads (three people), groups, subgroups and networks.Each SNA method or statistic views these components and their relations from a particular perspective.

There are commonly established social network measuresused to make structural patterns visible. Five common SNA statistics are:1) Socio-centric density, 2) Freeman betweenness centrality, 3) Network centralization, 4) Core-periphery analysis and 5) SNA cliques.

Socio-centric density & Freeman betweenness centrality are measures that consider individual nodes (people) in a network.This is sometimes referred to as the micro view of SNA because they are focused on members.Density calculations consider two types of network:value networks and dichotomized networks.In value networks tie strength is measured, while in dichotomized networks there is a binary "tie" or "no tie" indication.For both network types, density is highest in a network where everyone talks with everyone else, and lowest when there is little interaction among the network members. Freeman betweenness centrality (or simply, "betweenness") is a measure of the importance of a node to making connections between other nodes.High betweenness for a memberindicates that they are a "connection point" for ideas between two clusters within the larger group. For example, if you are the only person in your family who plays rugby and goes to church, you have a high betweenness in a network composed of a church and rugby team.

Network centralization, core-periphery and clique statistics analyze the entire social network and help to make visible thegroupings and relationships between groupings, sometimes called the macro view of SNA [3,4,6,18].Network centralization measures indicate how tightly the social network is organized around its most central point.Highly centralized networks have one big cluster. In our data, the semantics of these measures include every single event that occurs between members.High in-degree centralization indicates that group members are reading the posts made by a small subset of the class.High out-degree centralization means that a small number of students are making most of the posts.Core-periphery measures showthat there is some group in the core, and some other group in the periphery of a network with one center [25].Core members are distinguished by their connection with every other core member; they are influential.

A small number of studies have enriched incomplete, uni-directional trace data with participant interviews to create richer depictions of completely online learning. Reffay&Chanier [17] used member reports of activity and email logs from four experimental groups in a 10 week, online language course. They demonstrated the use of integrated core/periphery clusters and cliques by using core/periphery analysis to determine the optimum n for performing n-Clique analysis on a social network in an online course.Koku& Wellman [13], rather than using logs, used interviews from one point in time to look at betweenness in virtual communities. However, they did not apply this analysis to distinguish behaviors at the small group unit of analysis. These studies underscore the recognized importance of the analysis presented here, along with the difficulty past researchers have had obtaining the data needed to describe the structure and evolution of completely online groups.

Our research compares social networks visible from online interaction logs across three online activity types:Peer to peer activities, small group activities and individual activities.Using the previous network measures that incorporate member positions in a network along with measures of the network itself will make a more representativeonline social structure of these three different activity types visible.

3. METHODS

Through the use of bi-directional data logs, surveys and interviews, this study comprehensivelyexamines the social interaction patterns that occur in completely online groups over time, with each time period "snapshot" representing a particular activity type.Five completely online groups were studied as they proceeded through these individual, peer-to-peer, and small group activities.In this study we pursue these <u>three research questions</u>:1) How do the social interactions among completely online group members vary by activity type?2) How and to what extent do the structure and patterns we see in qualitatively analyzed data also have a visible representation in the sociograms?and 3) How do bi-directional data logs (as opposed to uni-directional data logs) contribute to the understanding of the social interactions and activity types seen in the analyses?

3.1 Study Context

This study was conducted at a large US university using an online learning environment composed of an open-source course management tool called Sakai and an open-source activity notification tool called CANS (Context-aware Activity Notification System), which logs bi-directional data. Students used Sakai wikis,JForum discussion boards integrated with Sakai, chat rooms and file storage areas during the course.The CANS logs includeda comprehensive record of the activity in Sakai including all read and post activity, the context of that activity (such as thread title or artifact name), and other members who created and read each discussion post and artifact.

3.2 Participants and Activities

Sixteen students participated in a semester-long, completely online course about designing CSCW software systems during a fall term.Students were randomly assigned into five groups at the beginning of the term and stayed assigned to those small groups for the duration of the study.

The course included 8 separate activities (each with its own activity type)that built upon each other, resulting in a series of design criticisms, individual interaction designs and group interaction designs.Each module also included an opening discussion question to which all members of the course contributed.Each of the eight activities was structured using one of three different a priori group structures:Individual, peer to peer and small group.The course was organized as follows:

Table 1 - Module Description, Length and Activity Type

Module Description	Length	Activity Type
Module 1: Getting started with CSCW.Main activities are reading and discussion board participation	2 weeks	Individual Activity
Module 2: Conceptual framework for CSCW and interaction design.Proposal for a better CSCW support system.	2 weeks	Individual Activity
Module 3: What is CSCW? Review examples and identify approaches.Critical analysis of existing CSCW Systems.	3 weeks	Small Group Activity
Module 4: Coordination Support.Prospectus for a CSCW system, review of prospectus with two partners.	2 weeks	Peer to Peer Activity
Module 5: Shared Workspaces.Analysis of Designs using user interviews.	2 weeks	Small Group Activity
Module 6: Sociality.Individually-developed visual prototypes and review of two buddies prototypes	2 weeks	Peer to Peer Activity
Module 7: Evaluation and understanding.Evaluation of prototypes.	2 weeks	Peer to Peer Activity
Module 8: Review and Reflection	1 week	Individual Activity

"Individual activity" was an activity not requiring group work. For "peer-to-peer activity", each course member had his or her work reviewed by another course member. For "small group activity", the small groups (assigned at the start of the term) needed to work collaboratively in order to complete the assigned task.

Bi-directional, contextualizedlogs of all participant interactions were recorded throughout the entire course.Students also received daily CANS email digests reporting fellow student read and post activity and information about new announcements and resources.

To triangulate our log data, three participants from three different groups were interviewed three times.Theseinterviews followedthe individual activity, peer-to-peer activity and small group activity respectively.Each interview lasted between 35 minutes and one hour.Informants who participated in interviews were recruited on a voluntary basis, paid ten US Dollars for each interview, and consented to their interviews being reported for research.All members of the course consented to analysis of discussion boards and course artifacts for research.

3.3 Analysis

For our analysis, a concurrent transformative mixed methods research design [6] was used. As shown in table 2, the social network analysis using the bi-directional logging data was triangulated with discussion board content, interviews, and artifacts using grounded theory techniques.

Table 2 - Data, Analysis Methods & Guiding Theories

Data	Analysis Methods
Bi-directional data logs	Social network analysis
Discussion board content	Content analysis, Grounded theory open coding & axial coding
Course artifacts and interview transcripts	Grounded theory open coding and axial coding

3.3.1 Network Analysis with Bi-Directional Logs

Social network analysis was used to address our first research question, which is how and to what extent interactions and group structure vary by activity type.Our interaction analysis was guided by structural theories of social organization[3,6,18],and constructed fromCANS data.Individual actions (reads, posts and responses), which occur mostly (92% of events) in the discussion board,are the core data that define network structure in our analysis.The discussion board we use is of the type where up to five posts are viewable on a page. If "reader 1" views a page in a discussion thread, a network "read" tie is created between that person and each person who has previously posted content to that page.Figure one describes this, with $p_1..p_n$ being a "poster", and $t_1.. t_n$ being a timestamp. Read actions in other parts of Sakai (8% of events) require a "click" and are recorded as each event occurs.

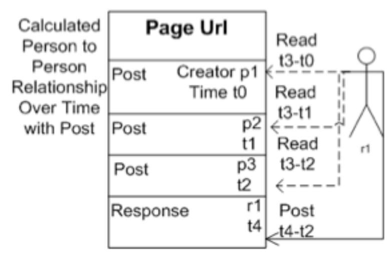

Figure 1 - Bidirectional Network Data: Reads and Posts

The integrated read and post behavior in our data is unique. We reference classic studies from the social networking literature and laboratory-based eye-tracking studies as a basis for recording discussion reads in this way. First, recordingthat a user "read" a discussion post simply because it was on a page they viewed follows the tradition of classic social networking studies, where presence at a party constituted a tie[7]. Second, eye-tracking studies suggest that users exhibit broad, top-to-bottomscanpaths and consistent fixation durations on web pages used for business (task) purposes, though more research is required to distinguish discussion board pages[16]. The results of this study and future studies will strengthen the theoretical foundation for recording log based, bi-directional network ties.

Our social network analysis followed strategies described by Scott [18], Wasserman & Faust [25] and Carrington et al [4]. We first built a valid social networking data set, and then selected which types of SNA most closely aligned with our research questions.Since our interests were group formation and development for different activity types within a group of 16, macro and micro social network analysis measures were pursued [6,18].UCINetwas used to perform the calculations, and netdraw to perform visualizations [3].UCINet provides a software implementation of network mathematical and graph theoretical constructs [3,6,18].

Density and centrality were examined as in deLaat et al[14].To build a more complete understanding of the network we incorporated the additional measures of betweenness, core-periphery analysis and clique formation, following the work of Reffay and Chanier[17].

Cleardistinctions among group members and activity types were first exposed through sociograms. Over 80% of discussion board events were “read” events. The completeness of this activity log data has implications for certain social network measures, like betweenness centrality. In our study, betweenness centrality is more meaningful than in prior studies examining online social networks because both reading and posting behavior between individuals was present in our data. After sociogramswere created, distinctions were subsequently brought into focus with statistical analyses and a description of the relationship between social network statistics and our qualitative findings on group experience, which we discuss in the next section.

3.3.2 Qualitative Analysis: Content & Interviews

Content analysis was used to address our second research question, which examineshow and to what extent participant experiences in each of the activity types is expressed in the sociograms.Discussion boards, interview transcripts, and other course artifacts were used to understand participant experiences.Theories of self-identification into groups and social structure were applied in our coding of discussion boards, interview transcripts and other course artifacts.To build an understanding of the differences in social behaviors across these activity types, we performed two types of analysis.

First, at the discussion board post unit of analysis, Tajfel’s[24] empirical and theoretical work on self-identification into groups was used to identify the formation and development of groups in different activity types. Two raters categorized communication as “Interpersonal”, “Interindividual”, “Intragroup” and “Intergroup”.Interpersonal communication occurs between two people.Interindividual communication also occurs between individuals, but those individuals are members of an identifiable group that includes others.Interindividual communication encompasses side conversations and behind the scenes negotiation.Intragroup communication is communication that occurs within a discernable group.Intergroup communication occurs between discernible groups.

Second, we applied methods from the tradition of grounded theory, beginning with open coding [23] to uncover themes and patterns from the data.A second pass at the open codes was performed using axial coding that was informed by the structure of both our initial open codes and our content analysis.This was done to look for data that helped to explain social phenomena that unfolded during each of the three activity types.

The three sets of participant interviews and daily researcher observation were used to triangulate findings from social network analysis and discussion board coding.We observed discussion board activity and assignments completed by each group at least once each day during the entire 16 weeks of the class.Field notes were maintained in nVivoand later analyzed alongside our other data.

For the final step in our process, a case study analysis method [22] was used.This addressed the unique challenges associated with integrating social network analysis and qualitative data.The case studies describe how participants experience online group development in each of the three activity types.Interview transcripts, field notes, course deliverables, course discussion boards and activity logs were coded as described above using nVivo 8.The 512 open codes were refined down to a core set of themes and patterns.

3.4 Summary

The combined use of qualitative analysis methods and social network analysis is intended to bring semantic richness to network data and bring structure to group trajectories within and between completely online collaborative activities.The comprehensiveness of our network data, combined with the breadth of source triangulation and manual coding of artifacts helped us to construct a uniquely rich view of completely online group work.

4. RESULTS

We learned that group structure varied in a clear way with activity type in this completely online course.The results are presented in two mainparts, corresponding with our research questions.First, we describe and contrast cases of group formation across three different a priori activity types:Individual, peer-to-peer and small group.Second, we integrate our qualitative analysis of group experiences with rich social network data culled from bi-directional activity logs and draw comparisons.The answer to our third research question, comparing bi-directional logs to uni-directional logs is woven into results for the first two questions and summarized at the end of the results section.

4.1 How Online Social Behaviors Vary by Activity Type

4.1.1 Network Density by Activity Type

The density information we present contrasts the same people and context across three time periods, each reflecting a different activity type.Density does vary by activity type.The course network is considerably denser for individually oriented activities than it is for small group and peer-to-peer activities.As shown in Table 3, individual activities have higher network density (8.43 to 3.67 or 3.28) than activities with a priori groupings.Network density is roughly the same for both small group and peer-to-peer activities.When we predetermine a network by defining groups of students (either small group or peer-to-peer) a priori, the interaction between individual nodes diminishes.If we tell people who to interact with, a substantial portion of the population interacts only with those people.

Two density measures presented in table three, binary and valued, help us understand these differences from two perspectives. The first perspective, binary, shows that roughly ¾ of the possible connections within the course are made at least once during the individual activity, while roughly ½ of them are made during small group or peer-to-peer activities.It is easier to compare binary measures across networks.The valued network scales up for individual activities on a steeper curve than the binary measure.This shows that not only are more connections made at least once during the individual activity than the others, but those connections occur with greater frequency.The strength of connection is higher. In other words, when completely online students in this course are not assigned a group to work with, their interactions with each other are more frequent and diverse.

In one sense, these results are intuitive.The finer-grained the a priori partitioning of the network is, the lower the overall network density becomes.This mirrors behavior we see in face-to-face groups. We meet more people at the bar than when we are seated at our assigned table.If, however, we consider that the cost of interacting with people other than those you are assigned to a priori is much lower in a technology oriented mode of interaction, it is nominally surprising that these kinds of observable patterns from the face to face world persist.

Table 3 - Network Density by Activity Type

Activity Type	Binary	Valued
Individual	0.7251	8.4357
Small Group	0.5351	3.6725
Peer-to-Peer	0.4963	3.2868

4.1.2 Network Centralization by Activity Type

Our nominally perplexing density data takes on additional color when integrated with analysis of network centralization across time and activity type (see table four).Network centralization measures show how tightly the social network is organized around its most central point.Our bi-directional logs allow us to produce a sharper measurement of in degree and out degree centralization because we have logs of both read and post activity.In this study, in degree centralization is a measure of the number of people who read or respond to a message from a course member.Out degree centralization is a measure of how many different people a particular course member has posted messages in reply to or read.High in degree centralization in this network indicates that activity in the course is focused on the posts of a small subset of its members.High out degree centralization in this network indicates that a small subset of the course is making the majority of posts.

Table 4 - Network Centralization Measures

Activity Type (three different time periods)	In degree Centralization	Out degree Centralization
Individual	42.65%	11.33%
Small Group	16.40%	11.35%
Peer-to-Peer	37.82%	9.63%

Our trace data shows low centralization in posting behavior (out degree) across all three activity types; in other words, there is no central group of people making the majority of the posts.In fact, it hardly varies at all.However, there is much greater centralization in the course network with regards to whose work is read (in degree), meaning there are some core members whom many others are learning through peripheral participation.Unlike network density, where the SNA measurements follow what we might intuit from the a priori structure, with network centralization we see our greatest structural distinction occurring with read activity (in degree) during small group activities.With both “individual” and “peer to peer” activities there is a cluster of “readership power” in the course, measured as in degree centralization.Some posts draw more attention than others for a significant number of participants (instructor activity is excluded from our analysis).In the small group activity, members stay focused on the other members of their group, and centralization at the course network level is relatively low (compared with the other activity types in this course).

A view of the density statistics along side the centralization statistics for the members of this course demonstrates that individual activities provide both the densest and most centralized network.The difference in centralization between the individual and peer-to-peer activities is not as great as their differences in density, however.In degree centralization statistics highlight an important aspect of online group work that is not clear from the density measure alone:In these online small groups, members seldom looked outside their group, even though the marginal cost for doing so is small.The density differences do not make this clear, because they are muted by the rough equivalence in post behavior.The in degree and out degree centralization measures parse member interactions in a finer way, which make important differences in the nature of small group activities in this online course visible.

4.1.3 Core-Periphery Analysis

Core members remain consistent across time and activity type in this course.Members of the core in core-periphery analysis are distinguished by having a connection to each other core member.Core members form what is known as a complete graph with each other.Periphery members, in contrast, are connected to a partial set of the core members.In our data set, core andperiphery measures across the whole community (16 people) are evenly distributed for individual and peer-to-peer activities, but centered on a smaller set of people in the small group activity.Participants whose activity puts them in the core for small group work also appeared in the core for other types of work.The core group remains the same, regardless of activity type.This is consistent with the findings of Reffay and Chanier[17], and confirms the limited utility of this type of analysis for distinguishing between activity types in an online course.

4.1.4 Freeman BetweennessCentrality

Freeman betweenness centrality, or simply betweenness, is a measure of the importance of a node to making connections between other nodes.In this network, a person could have high betweenness if they were reading or contributing to discussions across or “between” multiple different subgroups.This indicates the member is a “connection point” for ideas between two clusters within the larger group.The betweenness data in table fiveshow a wide range of values for individuals across activity types.

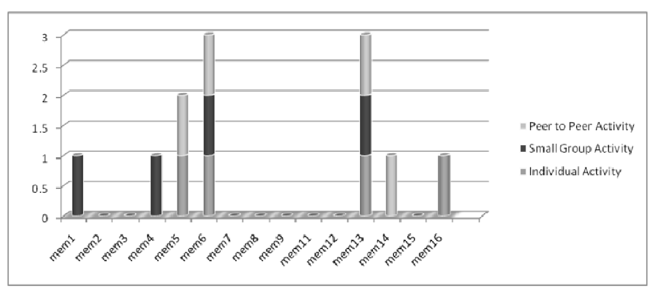

Figure 2- Highest BetweennessMembers by Activity Type

Table fiveshows that there is a small set of up to 1/3 of participants (e.g., mem13) with high betweenness for each activity type.Membership in this top 1/3 of betweenness differs for each activity.Figure two shows that seven different members are in the top quartile of betweenness for at least one activity, and that five of those seven are not in the top quartile for all activities.The implications of the diversity of member behavior across activity types will be clearer as we integrate the SNA data with our qualitative observations.The identification of participants with high betweenness establishes who within the course is actively spanning different discussion topics and different groups of others.

Table 5 - Individual Betweenness by Activity Type

Person	Activity Type (three different time periods)		
	Individual	Small Group	Peer to Peer
mem13	10.37	11.263	16.728
mem4	4.685	11.394	3.000
mem11	1.619	0.979	3.378
mem12	0.558	7.704	4.263
mem1	2.984	20.501	0.167
mem7	3.791	1.312	3.594
mem14	0	3.818	16.899
mem8	2.824	0.175	4.583
mem15	3.058	1.801	1.117
mem16	21.135	2.838	2.567
mem5	10.37	9.525	11.618
mem2	2.739	0.658	10.804
mem3	0.904	5.731	7.600
mem9	0.625	0.675	0.478
mem6	7.227	14.147	16.516

Memberswith high betweenness centrality are connectors between small groups, and in traditional social network analysis these membersare in a position of influence with the communities they span.The inclusion of passive participation behavior and active participation behavior changes, to some degree, the meaning of the betweenness statistic with our data set.Members whose betweenness is high, based on read activity, are engaging in observational behaviors more closely associated with legitimate peripheral participation than influence.

If we integrate our understanding of in degree centralization (read centralization) and betweenness, an interesting picture emerges. First, a small set of members is making the most interesting posts. With bi-directional logs, these members will have “high betweenness”, suggesting“thought leadership”. These thought leaders remain invisible using currently prevalent tools.Making these individuals visible to themselves, other course members and the instructor has potential for increasing participation by others and providing the instructor valuable understanding.

4.2 Group Experience and Network Structure Viewable Through Sociograms

In this section we examine how the different social network structures available in our data are visualized using sociograms.We then triangulate the sociograms with qualitative data. We do this in two steps.First, we describe cliques, which are visualized easily in sociograms.They show in a clear, visual way that the structure of the course members corresponds with activity type.Second, we connect the sociograms to our qualitative data and additional social network statistics.This triangulates our social network data and adds depth to our understanding of the completely online group interactions that take place during different activity types.

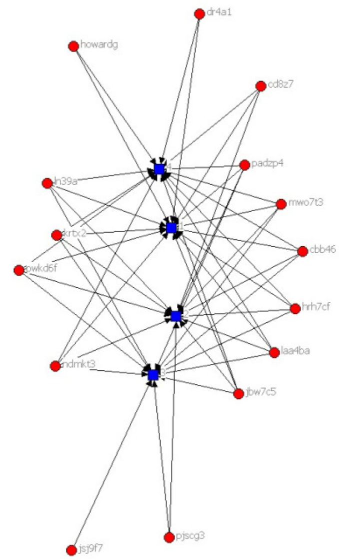

Figure 3 - Individual Activity Cliques (time 1)

4.2.1 Cliques

The social networking definition of a clique is a group of people who are connected to each other at some degree of indirectness.In the case of our analysis, we identified 1-cliques, which mean that the individuals in our cliques are connected directly with each other.In a small group of 16 participants, going beyond the 1-clique measure (2-cliques would be next) results in an analysis that presents the entire course as one big clique.Even in a large network, the number of hops to get to any individual is relatively small.For example, it is a party trick of legend to note that every actor in Hollywood is within Kevin Bacon’s 7-clique.Most are within Kevin’s 2-clique (aka, a “Bacon Number”).Even Brett Favre, a prominent American Football player who appeared in one movie, There’s Something About Mary, has a “Bacon Number” of 2.

During individual activities four distinct 1-cliques emerged, with some members participating in more than one clique.This corresponded with interview transcripts where members described small groups emerging during individual activities.The membership identified in these cliques is the same as the membership described in interviews.Figure three shows four cliques with overlapping membership form during individual activities.Figure four shows that eight cliques with fewer members in each clique are visible during small group activities.Each assigned group was found together in at least one clique, though others (usually individuals assigned reviewing roles for that group) were also found within those cliques.

The successful identification of a priori subgroups using clique analysis suggests that subgroups in general are discernible from analysis of CANS logs.This is what we would expect to occur during group relations in physical space.If you assign people to smaller groups, they are likely to engage with those groups, and possibly one other.If the goal is individual, a small number of groups may become visible through observation.Our coding and analysis of discussion posts does not show all of the same groups identified by SNA. This reinforces the power of bi-directional logs for revealing latent structure.

4.2.2 SociogramsRevealing Network Structure

First, as the clique analysis shows, there are differences in the structure of groups that are viewable through thesociograms.The differences are illustrated with clarity in figures three and four.

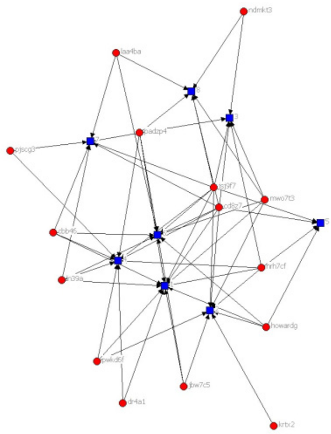

Figure 4 - Small Group Activity Cliques (time 2)

The comparison shows that when members are assigned to group work,this is visible in the sociograms.We can also see which members have an interest in their own group, and other groups.In an online learning situation, the instructor might value these insights.For example, we can see that some members in the sociogram in figure 4 are members of one or two groups, while other members participate in as many as five different groups.The visual structure of the three activity types triangulates our density data.

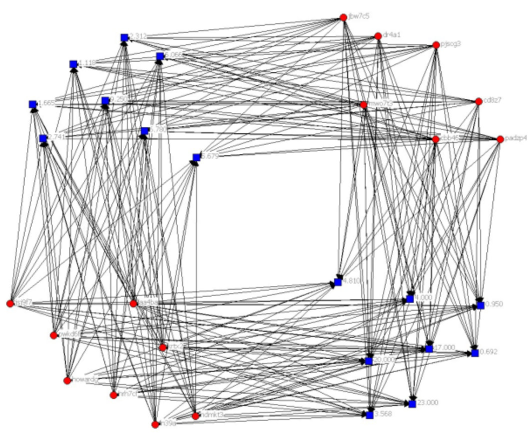

Figure 5 - Peer-to-Peer Activity Cliques (time 3)

The peer-to-peer activity (see figure 5) provided the largest number of cliques (15 total). This reflects the structure of the activity nearly exactly.Each member peer reviewed another course member.A different course member than the one whom they reviewed, then reviewed them, in turn.Variance of cliques from an expectation of the precise, a priori arrangement is near zero in the peer-to-peer activity.

4.2.2.1 Influence of Bi-directional Data

Analysis of only the post (write) interactions does not produce meaningful sociograms.In the interest of space, we will not display these ill-formed images.In the case of the individual activities, there is no clear group formation from the post-only logs.In the small group activities, we do see the a priori groups, but the other, "side groups" we see in figure 4 do not emerge.In the peer-to-peer activities, we see 14 of the 15 cliques emerge with post only data.Our analysis, then, is that the ability to understand differences ingroup structure during different activities in a visual way may be significantly enhanced by the use of bi-directional log data. More study is required to confirm this.

4.2.3 Betweenness& Qualitative Data

A second important finding is the relationship between these visualizations, the qualitative data and the social network statistic of betweenness centrality (or betweenness).There are two types of high betweenness members: lurkers and group creators. In the first type, high betweenness members werelurkers, or those with a high number of reads and a small number of posts. These members participated on the periphery and were invisible to the instructor and others with current tools. The second type of high betweenness members, group creators, emerged during our analysis of individual activities. The cliques identified through SNA during individual activities significantly triangulated with the groups identified by our interviews, and to a lesser extent with the ad hoc groups identified through discussion board coding.We noticed that, at this same time, user "mem16" had very high betweenness centrality. Mem16 was an active reader and poster during individual activities. Unlike high betweenness members who participated peripherally (read but didn't post), Mem16's activities led to ad-hoc group formation; Mem16 was a "group creator." After noting this trend, we went back and identified two key characteristics of Mem16's posts. First, the posts were provocative and often led to intense discussion among other members (a higher number of posts). Second, Mem16's posting behavior spanned more discussion threads than any other member.

4.2.4 Core-Periphery& Qualitative Data

When talking with informants and analyzing discussion boards using content analysis, a set of members emerged as major contributors during each of the three activity types.Our measurement of contribution level emerged from a combination of data analysis methods.First, we used content analysis of self-identification into groups and the method of constant comparison analysis from grounded theory as indications of group performance in small group activities.Second, our open coding of discussion boards and interviews was refined into a set of themes indicating levels of interactivity.Highly interactive themes from the open coding are associated with higher performing members.

Table 6 - Top Betweenness for Individual Activities

Individual	Betweenness
mem16	21.135
mem13	10.37
mem5	10.37

The members we identified as high performers also showed up as central figures in a number ofad hoc small groups.In addition, these members were present in the "core" of our core-periphery analysis.All except one of the members of the two groups who dominated our open coding for collaborative and supportive behavior are also in the core.The group member, who was in one of the top two performing groups, but not in the core, had the most connections with other members among those in the periphery.A high level of participation and diverse participation across activity types is associated with core membership in this online course.

Table 7 - Individuals in the Core are Clustered in Two Groups

Person	Activity Type			
	Ind.	Group	P-to-P	Group ID
mem1	core			1
mem2				1
mem3				1
mem4	core	core	core	2
mem5	core	core		2
mem6	core	core	core	2
mem7				3
mem8				3
mem9				3
mem10	core			4
mem11			core	4
mem12				4
mem13	core	core	core	5
mem14		core	core	5
mem15	core			5
mem16	core		core	5

The a priori groups with the highest core membership (a priori groups 2 & 5) also have the fewest issues in group formation and development, as discerned from our open coding and content analysis for self-identification into groups.The other a priori groups, at different times, experienced disputes, missing team members, or an inability to understand and complete their work with a high degree of quality.Future studies should examine the relationship between core-periphery measures, group performance and conflict.Our work suggests that such a relationship exists, but it is not clear how it operates.

Table six summarizes the core and periphery membership in this course, highlighting the core nature of most members of groups two and five.Periphery membership is indicated by the blank column and row intersection.This core-periphery analysis shows that members were in the core or periphery across all activity types.

4.2.5 Centralization and Qualitative Data

Individuals who are central to the overall course network, as measured by in degree normalized centralization (more people read their posts and/or responded) were also coded as the 'coordinating person' for the highest functioning groups in the a priori group activity.Table seven shows the three individuals with the highest in degree centralization in the course overall.Mem6 is the coordinating member for group two, mem5 is another member of group two and mem13 is the coordinating member of group five.Recall that groups two and five are the two groups who built the strongest group identity according to our open coding and content analysis for social identity.

It is likely that the most central member in a network will also have complete graphs with the other most active members. This relationship between core membership and centralization is somewhat intuitive, and it is the definition of core membership in the core-periphery analysis. The coordinating role of the most central members is confirmedthrough open coding, content analysis and social network analysis. This reinforces the intuitive idea that key group members may be identified by their behavior. More powerfully, it shows that these members can be identified through analysis of bi-directional interaction logs.

Table 8–SNA Statsistics in High Performing Team

Person	In degree centralization	Open Coded Group Leader	Group Number
mem6	19.444	Y	2
mem5	7.683		2
mem13	5.319	Y	5

4.2.5.1 Influence of Bi-directional Data

Participation in the world involves listening and speaking.Face to face, leaders are not always the most vocal people in the room.The same holds true in our assessment of highly central members.Different members, whose leadership is not corroborated by our qualitative data, show up in analysis of uni-directional logs as leaders.Online, the loudest person does not lead.

4.3 Synopsis of the Importance of Bi-Directional Log Data to Our Analysis

Throughout our results we pointed out specific analysis types where the results with uni-directional (read) data are less relevant to what is actually happening in a completely online course.The analysis presented here is only possible with the type of bi-directional logging system that captures users, context information and interaction of all kinds.Standard weblogs do not identify groups clearly.

5. DISCUSSION

Activity type in a completely online course becomes visible in different ways through the application of social network statistics and visualizations.In this study we have comprehensively analyzed interactions in a single online course. We explained how network density, betweenness, and centralization from bi-directional log analysis vary by activity type in this course.Our results suggest the possibility of a new kind of awareness tool for online instructors or managers of highly distributed groups.Instructors might use network statistics and visualizations to monitoractivity and interaction patterns in their courses. Such tools could help them to ensure that the individual and group *experiences* are occurring as intended. Knowing what patterns to expect across activity types createsopportunities for instructors to take a closer look and, if necessary, intervene when there is deviation from the expected patterns.

There are also potential implications for student use. With the useof bi-directional log data, we are able to depict trends and patterns of member interaction. For example, students will be able to see not only where the "herd" is (indicating where they might need to be and what they might need to be doing), but also which members of the herd they travel with across different contexts within the online course. This has the potential to increase people's sense of social presence.

Our methods of understanding group structure over time advances prior online social network research.We analyzed data related to posts and data related to reads.This gives a more accurate and rich description of the online learning and interactions that are taking place. For example, we learned that members of the course are more densely networked when they are not assigned a priori groups.For "getting to know you" periods of the course, or any collaboration, it appeared that group work was less optimal.These differences were less clear when uni-directional data were used because we missed important "read connections." Members with high betweennessengaged in observational behaviors akin to legitimate peripheral participation, which Gurzick & Lutters conjectured as legitimizing the value of lurking[9]. Betweenness, calculated from our bi-directional logs, helped us understand who in a completely online environment was most actively and broadly trying to integrate into the course.The two types of betweenness members – lurkers and group-creators (mem16) – are revealed through our analysis and have potential to affect the experience of an online course. Coaching peripheral members toward the core and encouraging group-creators to remain active are normal, physical classroom strategies for encouraging group activity. In this paper we have identified how these members can be made visible in an online learning environment.

This new understanding about how group structure and experience can be made visible in a completely online learning environment has implications for instructional practice as well. Better awareness tools for instructors and students will help both to see the fullview of their environment,and subsequently gain more control over their previously veiled experiences.

6. ACKNOWLEDGEMENTS

The authors wish to thank David Gurzick, Rebecca Reynolds, Andrea Forte, Gerry Stahl and several anonymous reviewers for their feedback and input on earlier drafts of this paper.

7. REFERENCES

[1] Allen, I.E. and Seaman, J. 2007. "Online nation: Five years of growth in online learning", The Sloan Consortium, p. 1-25.

[2] Aviv, R., Erlich, Z., Ravid, G. and Geva, A. 2003. "Network analysis of knowledge construction in asynchronous learning networks", Journal of Asynchronous Learning Networks, vol. 7, p. 1-23.

[3] Borgatti, S.P., Everett, M. G. and Freeman, L. C. 2002. "Ucinet for Windows: Software for social network analysis", Harvard, MA: Analytic Technologies.

[4] Carrington, P.J., Scott, J. and Wasserman, S. 2005. Models and methods in social network analysis, Cambridge Univ Pr.

[5] Cress, U. 2008. "The need for considering multilevel analysis in CSCL research: An appeal for the use of more advanced statistical methods", International Journal of Computer-Supported Collaborative Learning, vol. 3, p. 69-84.

[6] Creswell, J.W., Plano Clark, V. L., Gutmann, M. L. and Hanson, W. E. 2003. "Advanced mixed methods research designs", Handbook of mixed methods in social and behavioral research, p. 209-240.

[7] Freeman, L. C. 2003. “Finding social groups: A meta-analysis of the southern women data”. Proceedings from Dynamic Social Network Modeling and Analysis: workshop summary and papers.

[8] Goggins, S. and Erdelez, S. 2010. "Collaborative information behavior in completely online groups", in Collaborative information behavior: User engagement and communication sharing, J. Foster, ed., ISI Global.

[9] Gurzick, D., &Lutters, W. G. 2009.Towards a design theory for online communities. Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology.

[10] Haythornthwaite, C. 2001. "Exploring multiplexity: social network structures in a computer-supported distance learning class." The information society, vol. 17, p. 211-226.

[11] Howison, J., Wiggins, A., and Crowston, K. under review. Validity Issues in the Use of Social Network Analysis for the Study of Online Communities.Journal of the Association of Information Systems.

[12] Katz, N., Lazer, D., Arrow, H. and Contractor, N. 2004. "Network theory and small groups", Small group research, vol. 35, p. 307.

[13] Koku, E.F. and Wellman, B. 2004. "Scholarly networks as learning communities: The case of TechNet", Designing for virtual communities in the service of learning, p. 299-337.

[14] deLaat, M., Lally, V., Lipponen, L. and Simons, R. J. 2007. "Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis", International Journal of Computer-Supported Collaborative Learning, vol. 2, p. 87-103.

[15] Lave, J. and Wenger, E. 1991. Situated learning: Legitimate peripheral participation, Cambridge Univ Pr.

[16] Pan, B., Hembrooke, H. A., Gay, G. K., Granka, L. A., Feusner, M. K., & Newman, J. K. 2004. The determinants of web page viewing behavior: an eye-tracking study. Proceedings of the 2004 symposium on eye tracking research & applications.

[17] Reffay, C. &Chanier, T. 2003. How social network analysis can help to measure cohesion in collaborative distance learning. In Designing for change in networked learning environments. Proceedings of the international conference on computer support for collaborative learning.

[18] Scott, J. 2000. Social network analysis : A handbook, London ; Thousands Oaks, Calif.: SAGE Publications.

[19] Shen, D., Nuankhieo, P., Huang, X., Amelung, C. and Laffey, J. 2008."Using social network analysis to understand sense of community in an online learning environment", Journal of Educational Computing Research, vol. 39, p. 17-36.

[20] Stahl, G. and Hesse, F. 2007. "Welcome to the future: ijCSCL volume 2", International Journal of Computer-Supported Collaborative Learning, vol. 2, p. 1-5.

[21] Stahl, G. 2006. Group cognition: Computer support for building collaborative knowledge, MIT Press Cambridge, MA.

[22] Stake, R.E. 1995. The art of case study research, Sage Publications, Inc.

[23] Strauss, A.L. and Corgin, J. M. 1998. Basics of qualitative research: Techniques and procedures for developing grounded theory, Sage Publications, Inc.

[24] Tajfel, H. 1982. Social identity and intergroup relations, Cambridge University Press.

[25] Wasserman, S. and Faust, K.. 2009. Social network analysis: Methods and applications, Cambridge Univ Press.