Social Network Analysis in HCI

Derek L. Hansen and Marc A. Smith

Introduction

Social network analysis (SNA) is the systematic study of collections of social relationships, which consist of social actors implicitly or explicitly connected to one another. Social network analysts characterize the world as composed of *entities* (e.g., people, organizations, artifacts, nodes, vertices) that are joined together by *relationships* (e.g., ties, associations, exchanges, memberships, links, edges). SNA focuses on relational data about what transpires between entities in contrast to attribute data about individuals. Network analysts focus on the patterns generated within collections of many connections. For individuals, SNA is more about "who you know" than "what you know" or "who you are." At the group level, SNA illuminates how each person's individual connections aggregate to form emergent macrostructures like densely connected subgroups. Using the mathematics of graph theory, social network analysts calculate and visualize the properties of networks and the social actors that inhabit them.

HCI seeks to improve the ways people interact with information systems, many of which support interactions between people. SNA can be applied in many ways to HCI concerns, providing theory and methods for better understanding and evaluating the diffusion and impact of CSCW innovations like social media systems. Network analysis can be applied to capture the social structure of a user population before, during, and after new technologies are deployed. Network datasets can be used to measure changes in patterns of relationships and workflow that are not

D.L. Hansen (✉)
School of Technology, Brigham Young University, 265 Crabtree Bldg,
Provo, UT 84602, USA
e-mail: dlhansen@byu.edu

M.A. Smith
Social Media Research Foundation, 2617 Hallmark Drive, Belmont, CA 94002, USA
e-mail: marc@smrfoundation.org

visible in more common metrics like counts of users and rates of resource usage. A network perspective distinguishes between simple population growth and the development of important social structures within that population. The success of some systems may depend, for example, on attracting smaller populations of users who create a denser web of connections than systems that attract larger but more sparsely connected populations (see Ren & Kraut, in this volume). Attracting users in the first place is another HCI concern for which network methods can be useful. For example, SNA can help identify potential influencers who occupy strategic positions in existing networks who can recruit new users most effectively.

Social networks have formed for as long as people have interacted, traded, and engaged with one another. While social networks have existed long before the Internet, recent social networking services, such as Facebook and LinkedIn, support the creation of large, distributed, real-time social networks. When these services are used, they often generate data that is valuable for basic and applied research purposes. Prior to the widespread use of digital information systems, generating records of social interactions was challenging. In the era of pencil and paper data collection, datasets were often subjective, small, and time bound. Today, many legal, financial, educational, recreational, and personal communication systems generate the materials needed to analyze webs of human relations. Social networks are present in collections of e-mail, instant messaging, text messages, phone call logs, hyperlinks, message forum posts and replies, wiki page edits, tweets, "pins," video calls, multiplayer games, etc. These activities all generate network data that can be captured at a scale and pace never before possible, opening up new opportunities in computational social science (Kleinberg, 2008). Network analysis of online interactions is also proving to be a new source of actionable insights for community administrators, marketers, and designers of CSCW systems (Hansen, Shneiderman, & Smith, 2010).

Social media network maps can be a useful way to create a higher level understanding of collections of messages and the connections among authors that form in many information systems. Network maps can reveal divisions between subgroups of users that would otherwise be difficult to perceive. Network metrics can also be calculated for each participant to highlight the few people in key locations in a population, such as network hubs or bridge spanners. Visualization of networks along with calculated metrics can provide useful illustrations and summaries of the shape of a connected population. For example, Fig. 1 shows a network created from the connections among Twitter users discussing "global warming."

The graph represents a network of 415 Twitter users whose tweets contained "global warming." There is a green edge for each follows relationship. There is a blue edge for each "replies-to" or "mentions" relationship in a tweet. The tweets were made over the 4-h, 54-min period from Sunday, 11 November 2012, at 13:46 UTC to Sunday, 11 November 2012, at 18:41 UTC. The graph's vertices were grouped by cluster using the Clauset–Newman–Moore cluster algorithm. Each group is presented alone in a box, separated from all other clusters. The graph was laid out using the Harel–Koren Fast Multiscale layout algorithm. The vertex sizes are based on follower values. Visual attributes of this network map display multiple facets of each user and their connections. The size of nodes highlights important people,

Social media network connections among Twitter users
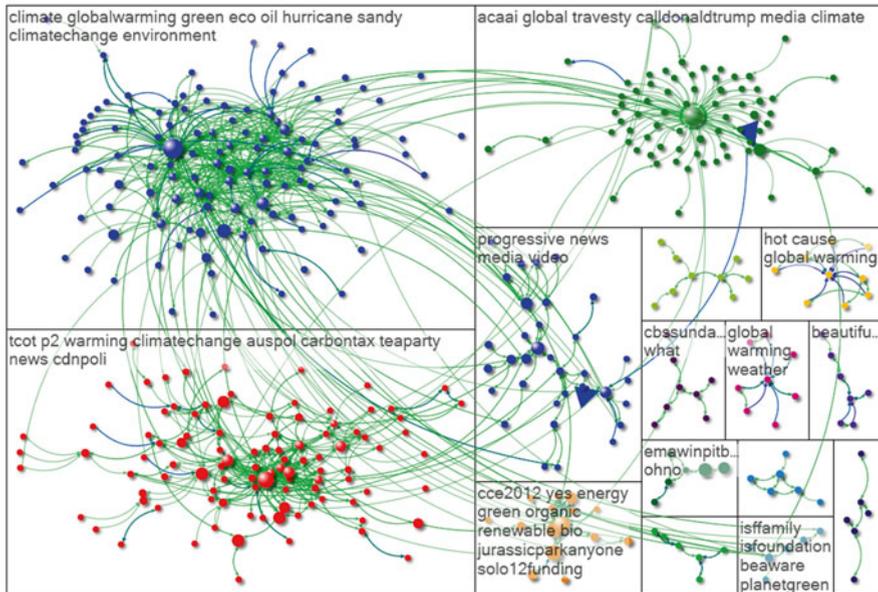
Created with NodeXL (http://nodexl.codeplex.com) from the Social Media Research Foundation (http://www.smrfoundation.org)

Fig. 1 A social network consisting of Twitter users (*nodes*) who have tweeted the word "global warming" connected to one another based on Follow, Reply, or Mention relationships (*edges*). Nodes are assigned different colors based on clusters, and hubs with many followers are indicated by size. Labels for each group are derived from frequently mentioned hashtags in the tweets from the users in each cluster. (Color figure online)

while color indicates membership in subgroups that are more densely connected to themselves than to other groups of users. The network is composed of a large connected component of people who are linked together by replying to or following one another. The connected group is subdivided into clusters or subgroups based on relative densities of connections. From analysis of this network and the content associated with it the groups can be labeled to indicate their focus or orientation. In this network climate change deniers are separate from people discussing climate science, sharing few follows, replying, or mentioning connections between the two groups.

A Brief History of Social Network Analysis

Though social networks are primordial, SNA is a relatively recently developed methodology whose history can be divided into roughly three phases: the foundational phase, the computational phase, and the network data deluge phase. See Linton Freeman's book on the development of SNA for a full treatment of the history of SNA (2004).

The early foundational phase, beginning in the eighteenth century and continuing into the 1970s, focused largely on defining terms and establishing the necessary mathematical graph theory foundation. Very early work by the famous mathematician Leonhard Euler demonstrated the value of using a graph theory representation to solve mathematical puzzles. In the 1950s and 1960s work by Paul Erdős and Alfréd Rényi provided formal mechanisms for generating random graphs that made statistical tests of network properties viable. Meanwhile, sociologists including Auguste Comte and Georg Simmel saw patterns of social ties as the main focus of sociology in contrast to the study of individuals and their attributes. During the 1930s, authors including Jacob Moreno, Lloyd Warner, and Elton Mayo applied formal mathematical methods to describe, analyze, and visualize networks in what was then described using terms such as "psychological geography," "sociometrics," and "sociograms." Stanley Milgram, working in the 1960s, performed his famous "six degrees of separation" study involving chain letters sent across the United States from random people to a stock broker in Massachusetts (1967). The average number of people needed to complete the chains was six, a surprisingly low number that illustrated how closely connected two individuals can be, even in extremely large social networks. In the 1970s, sociologist Mark Granovetter demonstrated the value of a social network approach by showing that "weak ties" (e.g., connections to acquaintances) were a much better source of new jobs than "strong ties" (e.g., family and very close friends) (1973). Later studies showed the "strength of weak ties" in other contexts including learning novel information, marketing, and politics.

The second major phase of the development of SNA, occurring largely in the 1970s through the mid-1990s, included the creation and systematic use of computational tools and methods. SNA as a methodological approach came into being during this phase, which leveraged the new capabilities of computers to analyze and visualize networks in novel ways. Lin Freeman built early tools for exploring networks (e.g., UCINet along with Borgatti and Everett) as well as identified core "centrality metrics" that provided objective measures of an individual's importance in a given network as described later in this chapter. George Homans developed new techniques for identifying subgroups (i.e., clusters) in networks, while Harrison White developed techniques for finding people that occupy similar network positions (via "structural equivalence"). Sociologist Barry Wellman founded the International Network for Social Network Analysis in 1976, which has served as a hub for social network researchers in a variety of fields ever since. Wellman has argued that SNA is not simply a method but is the core paradigm for explaining social action, particularly in our age of "networked individualism" where our work, community, and familial relationships no longer fit nicely within densely connected and bounded groups (2001). By the mid-1990s SNA was a well-respected approach in numerous fields ranging from organizational behavior (e.g., work by Ronald Burt and Rob Cross) to social psychology (e.g., Alex Bavelas' work) to communication networks (e.g., Noshir Contractor's work) to epidemiology. Perhaps the culminating work of this era is the "SNA bible" *Social Network Analysis: Methods and Applications* by Stanley Wasserman and Katherine Faust (1994), which rigorously summarized decades of research into a coherent mathematical framework, identifying the core metrics and techniques used by SNA tools and researchers today.

The current phase of SNA centers around the deluge of rich network data captured at Internet scale. A wealth of real-time social network data is captured by our everyday use of mobile phones, social networking sites, and commercial transactions. No longer is SNA a purely academic exercise, as corporations, governments, and nonprofit organizations utilize SNA techniques to find criminals, rank Web sites, recommend books, identify influencers, and restructure organizations. Authors such as Lada Adamic, Albert-László Barabási, Bernardo A. Huberman, Jon Kleinberg, Mark Newman, Steven Strogatz, and Duncan Watts have identified theoretical models that explain network generation and dynamics (e.g., see Newman, Barabási, & Watts, 2006; Newman, 2010), shown how information and influence propagate through them, and developed techniques for identifying clusters (i.e., communities) within them. Meanwhile, tools such as Pajek, developed by Vladimir Bagatteli, and the Stanford Network Analysis Platform (SNAP) by Jure Leskovec allow analysis of social networks at a scale never before possible. Other tools such as NodeXL and Gephi have focused on supporting SNA novices in their attempts to visualize small- to medium-sized networks. Computational social scientists have seized the moment by mining data from Facebook, Instant Messaging services, and other social media channels to more rigorously substantiate earlier work such as Milgram's 6 degrees of separation study, as described later in the chapter. Meanwhile, Nathan Eagle, Alex (Sandy) Pentland, and David Lazer have pioneered techniques for inferring friendship networks from data captured via mobile devices (2009). No doubt, this phase of SNA will continue to flourish as our social lives become increasingly mediated by technology.

Social Network Analysis and Human–Computer Interaction

Network analysis is a relatively new methodological and theoretical framework used within the HCI tradition. However, it has become prevalent in recent years, as social technologies have blossomed and tools for analyzing and visualizing networks have become more widely available. In this chapter we focus on how SNA can be used to design, evaluate, and understand CSCW and social media systems. We begin by describing five different goals that HCI researchers and practitioners can use SNA to achieve. We then move on to a discussion of specific questions that SNA can effectively address.

Goals of Social Network Analysis for HCI Researchers and Practitioners

1. Inform the design and implementation of new CSCW systems.

SNA can characterize the social structure of a population of intended users of a new CSCW system before the system is put in place. Understanding the social

network properties of a target user population can help clarify requirements and challenges, leading to better initial designs and implementation strategies. Research has shown that mapping the social network of members of a large organization can help design social and technical strategies to facilitate more effective information flow (e.g., Cross, Parker, Prusak, & Borgatti, 2001). For example, tools may be needed that identify important bridge spanners or encourage the increased connection of groups that are too disconnected. Those implementing a new CSCW system could use SNA to identify, educate, and leverage those who will influence the maximal spread of adoption through the network to assure its rapid, effective use (Kempe, Kleinberg, & Tardos, É, 2003) or help others to know how to use a new technology (Eveland, Blanchard, Brown, & Mattocks, 1994).

Data for these analyses may come from network surveys (Marsden, 2005) or from existing data sources such as communication exchanges (e.g., e-mail, phone logs, IM, texts). Networks from these sources can characterize existing social structures and establish a baseline for measures of the impact of new CSCW systems (Goal #3). Furthermore, individuals with unique and important network positions can be identified and interviewed or observed as part of a comprehensive contextual inquiry process (Beyer & Holtzblatt, 1997).

2. Understand and improve current CSCW systems.

SNA of data from existing CSCW systems can illustrate the ways current features are utilized by users in different locations in the network. For example, the phenomenon of "unfollowing" someone on Twitter is partly explained by the social network structures of those involved (Kivran-Swaine, Govindan, & Naaman, 2011). Basic understanding of the pattern of user interactions can often inform the future design of social and technical improvements to CSCW systems. For example, network analysis of a technical support message board forum can help identify those who fill vital roles, such as "Answer Person" (Welser, Gleave, Fisher, & Smith, 2007). Community administrators can court these people to encourage them to remain active.

SNA may help community managers understand what is happening in large-scale communities where reading through even a meaningful sample of the content is not feasible. For example, a subgroup of users labeled "Theorists" was identified using network analysis techniques from among hundreds of thousands of Lostpedia wiki editors (Welser, Underwood, Cosley, Hansen, & Black, 2010). Knowing this subgroup exists could allow designers to develop tools that meet the particular needs of subpopulations, such as page templates that help systematically compare the competing theories. Similarly, unique social structures were found in Wikipedia's "breaking news" articles, which lead to insights about how people coordinate and potential designs to improve such work (Keegan, Gergle, & Contractor, 2012). Recently, several studies have developed recommendations for improving virtual reality games based on network analysis of guild networks and social interaction patterns (Ducheneaut, Yee, Nickell, & Moore, 2006; 2007). Other studies have shown variations in network structure by different users (e.g., teens and older adults) of the same discussion forum software (Zaphiris & Sarwar, 2006). Network

methods that identify subpopulations can offer customized interfaces and services to different groups of users, using the history of other users in the same group as a guide. Education researchers have shown how students use different social features to interact within small groups and class-wide, with implications for system design and instructional strategies (Haythornthwaite, 2001). Work that shows separation between subgroups (e.g., conservative and liberal bloggers or readers) (Adamic & Glance, 2005) could be used to design tools that recommend posts that would increase cross-pollination of ideas (Munson & Resnick, 2010).

3. Evaluate the impact of CSCW system on social relationships.

SNA can be used to evaluate the impact of a CSCW system on the existing social structure of a population. Many CSCW systems are designed to, at least in part, influence the social relationships of those who use the system. Corporate intranets help employees find internal experts; online exchange markets match buyers and sellers; online community sites hope to develop sustainable communities around their niche topic; and collaboratories aim to facilitate scientific collaboration. Measuring the changes in aggregate and person-specific network metrics can help systematically evaluate the effectiveness of such systems. For example, the impact of CSCW systems designed to maintain weak ties between dispersed occupational communities could be measured (Pickering & King, 1992). Indeed, increased use of an internal, corporate social networking site has been shown to be positively associated with bonding relationships, sense of corporate citizenship, interest in connecting globally, and access to people and expertise (Steinfield, DiMicco, Ellison & Lampe, 2009). Evaluation can also be performed to assess the impact of a specific feature or social intervention. For example, the impact of an online "icebreaker" activity could be assessed by looking at changes in the network (e.g., network density) before and after. The majority of work in this arena relates to structuring social networks within organizations to improve knowledge creation, sharing, and innovation (e.g., Cross, Parker, & Borgatti, 2002; Borgatti & Foster, 2003; Müller-Prothmann, 2006). However, education researchers are also using network data to identify students using online course management systems that may be in need of extra support (Dawson, 2010).

Data for evaluation assessments may come from offline network surveys, existing communications (e.g., e-mail) captured over time, or system usage data (e.g., friendship or follow relationships). For large-scale evaluations, SNA can be used as part of a mixed method approach. For example, SNA can be used to identify individuals to interview based on their network positions (e.g., those with high, medium, and low network centrality; those from different subgroups).

4. Design novel CSCW systems and features using SNA methods.

SNA can be used as input to new CSCW systems and features. A growing number of research prototypes and innovative products leverage SNA metrics and methods to provide enhanced functionality. For example, a tool that recommends potential friends on a social networking site can use network properties to help

identify likely candidates (Chen, Geyer, Dugan, Muller, & Guy, 2009). SNA has been used to help identify experts in technical support groups (Zhang, Ackerman, & Adamic, 2007) and organizations (Ehrlich, Lin, & Griffiths-Fisher, 2007; Perer & Guy, 2012), though early work showed that users often did not trust that their personal friends were the best experts (McDonald, 2003). Early work showed that social structure coupled with temporal patterns could be used to develop situated awareness tools (Fisher & Dourish, 2004). More recent work has used SNA to identify political tendencies of the followers of different news agencies on Twitter (Golbeck & Hansen, 2011), a technique that could be used for tools that personalize news or present alternative views. Tools have been developed that leverage network analysis and visualization to help gain insights into large datasets, such as published literature on a topic (Chau, Kittur, Hong, & Faloutsos, 2011). A novel feature that would show network diagrams of researchers who use similar queries in Citeseer has been proposed to help identify potential collaborators and research communities (Farooq, Ganoe, Carroll, & Giles, 2007). Recent work has explored the theoretical and practical design implications for promoting "social translucence" within directed social network systems, such as Twitter, where users can only see a portion of the social space, unlike chat rooms and discussion forums where everyone is visible to everyone else (Gilbert, 2012b). Related work has proposed novel information dissemination strategies that leverage social networking sites and semi-anonymity, such as "veiled viral marketing" (Hansen & Johnson, 2012). These examples give a flavor of the countless possible uses of SNA to enhance current CSCW systems, making this a particularly ripe area of research.

5. Answer fundamental social science questions.

Network analysis of data from CSCW system can be used to address fundamental questions about the nature of social relations. This research is part of the growing field of "computational social science," a set of techniques that use computational techniques to address core social science questions in novel ways. Because so much data is automatically captured via social media, they provide new opportunities to test hypotheses and theories at a much larger scale than previously possible. For example, Leskovec and Horvitz analyzed data from 180 million Microsoft Instant Messenger users finding an average path length of 6.6 between users, strikingly close to Milgram's original 6 degrees of separation work (2008). More recent work based on Facebook shows an average path length of just under five (Ugander, Karrer, Backstrom, & Marlow, 2011). Another example is a study of Facebook (Bakshy, Rosenn, Marlow & Adamic, 2012), which helped support and extend Granovetter's original work (1973) that showed the importance of weak ties. Other work predicts the strength of ties between individuals based on their social media interactions (Gilbert, 2012a; Gilbert & Karahalios, 2009) or mobile phone usage patterns (Eagle, Pentland, & Lazer, 2009). Such data can support further large-scale studies of social networks by reducing the need for raw data collection from users. Other studies are looking at the factors that lead to the sustained growth or death of online communities, such as the initial network structure (Kairam, Wang, & Leskovec, 2012).

Social Network Analysis Questions

SNA has been used to address a wide variety of questions in dozens of fields. While these questions vary considerably, they all share an emphasis on understanding social structures and how those structures influence outcomes of interest. SNA is designed to answer several types of specific questions as the categorized lists below illustrate.

Questions About Individual Social Actors

Often, network analysts are interested in identifying individuals who play an important, prominent, or unique role within a particular social network. Analysts use "centrality metrics" and "equivalence metrics" to address these questions. Some example questions include the following:

- Who are the most popular individuals in a network (e.g., network hubs)?
- Which individuals have the most influence?
- Who is a bridge spanner between different subgroups of users?
- If one is trying to disrupt a network, who should be removed?
- Are there different types of social actors that can be identified by unique network patterns? Who fills those social roles?

Questions About Overall Network Structure

Many questions relate to the overall structure of complete networks, such as the network of all Facebook users or all employees of an organization. Instead of focusing on the position of individuals within the network, these questions focus on the overall distribution. Analysts use "community detection algorithms" (i.e., network clustering algorithms) and a variety of "aggregate network metrics" to answer these questions. Some example questions include the following:

- How interconnected are a group of social actors (i.e., how dense is the network)?
- What is the distribution of individual network properties or social roles? For example, are there only a small percentage of "hubs" with a majority of "isolates"? Are there "enough" people that fill certain social roles?
- Are there subgroups of highly connected users (i.e., clusters, cliques)? If so, how many? And what is their relationship to one another? How do they differ from one another?
- What network properties or *motifs* (i.e., recurring network patterns) are related to social outcomes of interest? For example, what are the network structures of highly efficient groups, teams, businesses, and markets?

Questions About Network Dynamics and Flows

Other questions look at how networks change over time (i.e., network dynamics) or how information, objects, and attributes flow through networks (e.g., information diffusion, technology diffusion). Some example questions include the following:

- How do the structures of social relationship vary over time? For example, does the network become more interconnected or diffuse with use of a CSCW system?
- How does the importance of specific individuals, social roles, or clusters change over time? For example, does an intervention designed to bring separate subgroups together have the intended effect?
- How does information spread through a network (e.g., Twitter)? How can information propagation be catalyzed or minimized? What other attributes spread through a network?
- How does the use of new technologies spread through social networks? Who influences adoption of technology the most?

Performing Social Network Analysis

Despite the many types of analyses that can be performed, there is a common set of key steps including identifying the goals of the analysis, gathering data, and visualizing and analyzing the data using various network analysis software programs. This is a highly iterative process (Hansen et al., 2009). Analysts refine their goals after realizing the limitations of their datasets. Exploratory visualizations help identify the types of quantitative analyses that should be performed. And, additional data is often needed to validate or refute preliminary results.

Identify Goals and Research Questions

HCI researchers use SNA to accomplish a variety of high-level goals, each of which includes a large number of potential subgoals and research questions. It is essential that analysts hone in on a few critical goals and turn them into specific research questions, lest they spend unreasonable amounts of time aimlessly meandering around the data. Having said that, within HCI, SNA is often exploratory in nature and as with some types of qualitative research, analysts may only recognize what they are looking for once they see it. Often, after a preliminary analysis of initial data the questions are refined, another round of data collection is completed, and a final analysis is performed.

Table 1 Key sources of network data

Data source	Comments	Effort level
Raw data from system usage (e.g., database or XML files)	If you have access to the source data for a CSCW system (e.g., you are hosting it), you can query the data directly in highly customizable ways	Medium–high
Network survey	Network surveys ask people to manually characterize their relationships with other people (e.g., "list (or select) the people you turn to most often for answers to technical questions"). These can be administered via paper or, more commonly, via specialized network survey software (e.g., Network Genie), which may generate lists of employees based on existing databases or manually entered names	High
Application programming interfaces (APIs)	Most major social Web sites such as Facebook, Twitter, and YouTube have APIs that allow programmers to request data. You may need to register first, and you will often be restricted to data they make available	Medium–high
Screen scraping	If APIs are not available for a site you can write custom screen scraping software or use existing tools (e.g., VOSON), though legal restrictions may apply (e.g., the site's privacy policy)	Medium–high
Network analysis importer tools	Some network analysis tools allow users to import data from third-party sites (e.g., Facebook, Twitter). These tools may show up as import wizards (NodeXL), plug-ins (Gephi), or stand-alone network data capture tools (NameGenWeb Facebook App)	Easy
Existing datasets	An increasing number of existing network datasets are being made freely available. Examples include the Enron e-mail network, Amazon-related items, and blog networks (see http://snap.stanford.edu/data/ for more)	Easy

Collect Data

The next step is to collect the data needed in order to achieve the desired goal or answer the designated research questions. Below is a description of the sources of network data, different types of network data, and ways of representing network data.

Sources of Network Data

Depending on the specific data needs, collecting data may take considerable effort or be as easy as checking the appropriate boxes in an import wizard of an SNA software tool such as NodeXL. Table 1 shows the key sources of data that can be used in network analysis. Those that require more effort typically allow for more flexibility in the specific types of data that are collected.

Types of Social Networks

There are many types of networks. The specific type of network will determine how to appropriately analyze, visualize, and interpret the data. The type of network is determined by the underlying phenomena it represents. For example, a network of Twitter Following relationships is different from a network of Facebook Friendships because Facebook friendships must be mutual (if you are my Friend I am necessarily your Friend), while Twitter follow relationships do not have to be mutual (I can Follow you without you Following me).

Below is a brief description of the key terminology used to characterize networks.

- *Directed Versus Undirected.* Directed networks represent phenomena where the connection between two nodes is not necessarily reciprocated. Examples include communication networks (e.g., I send you an e-mail; you reply to my forum post), exchange networks (e.g., I sell you something), and awareness or following networks (e.g., I follow your updates). Undirected networks are always mutual, for example, friendship networks (such as on Facebook where one cannot friend another person without their consent) and affiliation networks (e.g., we are connected because we are affiliated with the same organization or we both edit the same wiki page).
- *Weighted Versus Unweighted.* Some edges have values associated with them. For example, edges in an e-mail network are "weighted" based on the number of messages one person sends to another person, while a wiki coedit page network is weighted based on the number of pages two people have both edited. Other edges are binary; they either exist or they do not. For example, Facebook friendships and Twitter follow relationships do not have weights.
- *Multiplex Networks.* Multiplex networks include multiple types of edges. For example, a network that connects people together based on their communication via e-mail, phone, and face-to-face interactions would include three distinct types of edges. This could be analyzed and visualized as a single multiplex network or as three distinct networks.
- *Unimodal and Multimodal Networks.* Many social networks, called unimodal networks, include only one type of node. For example, all the nodes represent people. Or, all of the nodes represent organizations. In contrast, multimodal networks include more than one type of node. For example, a network may include people who are connected to organizations or another network may include people who are connected to wiki pages they have edited. If there are only two types of nodes we call the network bimodal, which is a subset of the more general multimodal concept. Many bimodal networks, called bipartite networks, have one type of node (i.e., people) connected to another type of node (e.g., organizations) without any edges connecting nodes of the same type (e.g., people to people). These bipartite networks can be transformed into unimodal networks. For example, the person-to-organization network can be transformed into a person-to-person network where people are connected by a weighted edge that

Edge List

Node1	Node2
Brian	Marc
Derek	Marc
Derek	Adia
Marc	Ben
Maren	Derek
Maren	Adia

Matrix

	Brian	Marc	Derek	Adia	Ben	Maren
Brian	0	1	0	0	0	0
Marc	1	0	1	0	1	0
Derek	0	1	0	1	0	1
Adia	0	0	1	0	0	1
Ben	0	1	0	0	0	0
Maren	0	0	1	1	0	0

Graph

Ben, Adia, Marc, Derek, Brian, Maren

Fig. 2 Three ways of representing network data

represents the number of organizations they are both a part of. Conversely, an organization-to-organization network could be created where a weighted edge represents the number of people who are part of both organizations.

- *Partial Networks.* In practice, it is not practical or useful to collect data on an entire network (e.g., all Facebook users). Instead, analysts create partial networks in a variety of ways. One approach is to create an "egocentric network," which includes a single node (called "ego") and all of the nodes that ego is directly connected to (called "alters"). When the connections between alters are also included, the graph is called a 1.5 degree network. Adding ego's "friends of friends" makes it a 2.0 degree network and so forth. Other techniques for creating partial networks include sampling a large network (Leskovec & Faloutsos, 2006) or finding some network boundary such as membership in an organization.

It is important to recognize that a single socio-technical system inevitably includes many types of networks. For example, Facebook includes the obvious friendship network (unimodal, unweighted, undirected), the "people tagged together" network (unimodal, weighted, undirected), the "wall post" network (unimodal, weighted, directed), and the "person-to-group" network (multimodal, unweighted, undirected) to name a few. The choice of which networks to focus on depends on the goals of the particular study.

Representing Network Data

Network data is represented in three primary ways: edge lists, matrices, and graphs (see Fig. 2). An "edge list," also called an "adjacency list," contains a row that represents each edge in the network. In directed networks the first column lists the "source" node and the second column lists the "destination" node. Additional columns can be used to describe the type of edge and/or weight of the edge. An adjacency matrix lists each node as a header for both the rows and the columns, with matrix values corresponding to the weights of the edge (or a 1 or a 0 if it is unweighted). Finally, a network graph visually shows the nodes as vertices (e.g., circles or other shapes) and the edges as lines connecting them. Visual attributes can be used to represent edge weights (line thickness or opacity), directionality (lines with arrows), and node types (different shapes).

In addition to the network data, additional attribute data that describes the nodes and/or edges is often included. For example, you may have data on each person's gender, age, organizational role, membership duration, etc. Network graphs can be customized to help understand how this attribute data maps onto the network. For example, larger nodes could represent online community members who have been around longer. An analysis may reveal that larger nodes are well connected with each other but not with smaller nodes (newer members).

In practice, there are several common network file formats that most network analysis tools can import and read. These include GraphML (.graphml), Pajek (.net), Graphlet (.gml), GraphViz (.dot), and standard text files (.txt or .csv).

Analyze and Visualize Data

A wide range of analysis techniques can be used to understand and characterize social structures. New network analysis methods, metrics, models, statistical techniques, and algorithms are developed by an ever-growing, highly prolific research community consisting of researchers from a variety of fields. In this section we introduce some of the most commonly used techniques, organized into a handful of major topics within network analysis. Readers looking for comprehensive coverage should look to the additional resources mentioned later in this chapter.

Network Analysis Tools

SNA requires the use of specialized software designed to compute network metrics and visualize network graphs. The tool landscape is in constant flux (see http:// en.wikipedia.org/wiki/Social_network_analysis_software for a comprehensive list). Table 2 describes five of the most commonly used tools in order of their sophistication.

Node-Specific Metrics: Focusing on the Trees

Analysts often want to characterize how important an individual is within a particular social network. Of course, there are many different ways that a person may be important. One person may be popular, another may serve as a bridge spanner between otherwise separate groups, and yet another may be connected to popular people despite having few connections of their own. Each of these is important in a different way.

Network analysts have developed a set of quantitative measures called "centrality metrics" to represent these various types of importance. The most commonly used centrality metrics are shown in Table 3. Several of them use the idea of the "distance" between two social actors, which is measured by the number of edges on

Table 2 Commonly used network analysis and visualization tools

SNA tool	Description	Expertise required	Open source	Maximum network size
Gephi	Stand-alone network analysis designed primarily for visualization. Can be extended via plug-ins	Designed for novices	Yes	Hundreds of thousands
NodeXL	Includes sophisticated graph visualizations, social media data importers, and extensibility via formulas and macros, but relatively few metrics	Microsoft Excel plug-in designed for SNA novices	Yes	Tens of thousands
Pajek	Includes a comprehensive list of network metrics and statistical tests. Steep learning curve	Designed for sophisticated analysis of large datasets	Yes	Millions
R	Open-source statistical package with social network analysis functionality via the igraph, sna, network, and statnet packages. Includes a comprehensive list of network metrics and statistical tests	Steep learning curve	Yes	Millions
UCINet	Includes a comprehensive list of network metrics and statistical tests. Designed for knowledge-able SNA researchers, but does not require coding	Designed for researchers performing social network analyses	No	Tens of thousands

the shortest path between two nodes (i.e., the geodesic distance). Variations of these metrics, as well as specialized versions of them appropriate for weighted and/or directed networks, are also available. These core metrics are calculated by all major network packages.

These metrics, along with statistical and visualization techniques, help identify the "structural signatures" of individual participants. For example, some users, such as news agencies on Twitter with their high in-degree, function as *network hubs* able to directly reach a large audience. Others may have relatively few followers on Twitter, except within a subset of users who discuss a certain topic (e.g., use the hashtag #CSCW2012), making them *topical hubs*. Users with high betweenness centrality often serve as *bridges* connecting otherwise disparate groups together by spanning "structural holes" (Burt, 1995). Users who are re-tweeted by several hubs will have high eigenvector centrality and may reveal individuals who serve as behind-the-scene *influencers*. Users who are not connected directly to others are referred to as *isolates*. Network analysts also differentiate between those in the *core* of the network (i.e., well-connected group at the "center" of the graph) and those on the *periphery* (i.e., the fringes).

Table 3 Common centrality metrics

Centrality metric	Description	Intuitive interpretation
Degree (in-degree and out-degree)	The number of edges connected to a node. For directed networks, the number of incoming links is the in-degree, while the number of outgoing links is the out-degree	Measures popularity (i.e., the number of friends one has). In-degree may measure the number of messages one receives, while out-degree may measure the number of messages sent
Betweenness	The number of shortest paths between all other nodes that a particular node is on—i.e., how often a node lies "between" other nodes	Measures how disrupted flows through a network would be if a person was removed. Helps identify "bridges spanners"
Closeness	The inverse of the average distance to all other nodes—i.e., how "close" a node is to other nodes	Measures how long it would take to disseminate information from a person to all others in the network
Eigenvector	A measure of a node's importance that considers the importance of the node's neighbors, where importance is calculated as a weighted sum of direct connections and indirect connections of every length	Measures not only the number of immediate connections (i.e., degree) but also the importance of the people one is connected to

At times it is helpful to identify classes of people who share a similar structural signature or position in a network. Such individuals often fulfill similar social roles. For example, Welser, Gleave, Fisher, and Smith used unique structural signatures to identify key individuals they called "answer people" within technical support Usenet newsgroups (Welser et al., 2007). These individuals have high out-degree (i.e., they answer many questions), are disproportionately tied to isolates (people with only one connection), and have few intense ties (i.e., multiple exchanges with the same person). Their initial insights gained from visualization were validated using regression analysis to predict and identify those filling this role (as identified through content analysis of messages) with high accuracy ($R2 = .72$). Another technique to identify social roles is to use *equivalence methods* to identify similar individuals based on their relation to others in the network (Wasserman & Faust, 1994). For example, employees all tied to a single manager and nobody else in the company likely play a similar professional role.

Aggregate Network Metrics: Focusing on the Forest

Network analysts have developed a language and set of metrics to help characterize the entire networks, just as they have to characterize the roles of individuals within those networks. This allows for the comparison of networks with one another or

Table 4 Common aggregate network metrics

Metric	Description	Intuitive interpretation
Density	Number of edges in the network divided by the number of possible edges	The amount of interconnectivity in a network
Diameter	Maximum geodesic distance (i.e., the longest "shortest" path) of all pairs in a network	The number of hops needed to reach two individuals who are as far away socially as possible
Average geodesic distance	The average geodesic distance of all pairs in a network	The average number of hops (i.e., degree of separation) between two people in the network
Network centralization	The sum of differences between the centrality of each node and the node with the highest centrality divided by the maximum possible sum of differences	A measure of how hierarchal a network is—i.e., how centralized it is around one or a few key social actors (where 0 describes a network where everyone is connected to everyone else and 1 describes a "star" network with one key person connecting everyone else together)

over time. Visualizing entire networks is often useful, as it can reveal overall structures such as the *core* or the *periphery* of a network, network *clusters* (see next section), and other patterns. However, many graphs are too large to meaningfully visualize and some properties of a graph are difficult to visualize (e.g., the longest geodesic distance) making the calculation of aggregate network metrics essential.

A different set of metrics help characterize the properties of an entire network. Like summary statistics (e.g., mean, standard deviation) help characterize attribute data, aggregate network metrics (e.g., density, diameter) help characterize network data. Also, like summary statistics, they only tell part of the story. Just as a mean does not provide any details about the distribution that generated it, a graph density metric does not provide any details about the network that generated it. Basic metrics include the number of vertices and edges, the number of connected components (i.e., clusters of vertices that are connected to each other through some path), and their size (measured in number of vertices). Other commonly used aggregate network metrics are shown in Table 4.

In addition to aggregate network metrics, network analysts often look at the distribution of node-specific metrics such as degree. This can help identify outliers and get an overall sense of the network. For example, a network that is centralized around a few key individuals, but otherwise not densely connected, will have a very skewed degree distribution with a couple of high-degree individuals and many very-low-degree individuals. In contrast, a more densely connected network where mostly everyone is interconnected will show a relatively constant (i.e., flat) degree distribution, since everyone will have a similar degree.

Network Clusters and Motifs: Focusing on the Thickets

Networks are composed of smaller components, which are often useful to examine in their own right. Some nodes may be highly interconnected forming a *clique* or a network *cluster* (see Fig. 1 for examples). Algorithms to identify these tightly knit groups are called many things including *community detection algorithms*, *network clustering algorithms*, *n-cliques*, *n-clans*, *k-plexes*, *k-cores*, *factions*, *blocks*, and *cut-points* (Hanneman & Riddle, 2005; Newman, 2010). Other recurring structures, sometimes called network *motifs*, show unique patterns such as *fans* (one person connected to otherwise isolated nodes), *tunnels* (nodes connected in a long independent chain), and *structural holes* (places where a lack of connections offers unique opportunities for those who span them) (Burt, 1995). At an even more granular level, triads (combinations of three nodes) serve as the building blocks of networks, inspiring network analysts to perform *triad censuses* wherein they characterize the distribution of the different types of triads (Hanneman & Riddle, 2005).

Many important insights can be gained from identifying and quantifying these network structures, since network topology often reflects social divides, political opinions, and other behavior of interest. For example, studies have shown a clear divide between liberal and conservative bloggers (Adamic & Glance, 2005) as well as distinct subgroups of Twitter users interested in gubernatorial elections from a national and local perspective (Himelboim, Hansen, & Bowser, 2012). Online community administrators can use network clusters to help identify potential conflicts and/or opportunities to bridge them. And system designers can identify how different collections of people utilize various collaborative features.

Network Dynamics and Information Flow

Thus far we have examined networks as static, unchanging entities. However, social networks are constantly evolving. Furthermore, information and other items can be distributed through networks over time, as happens in viral marketing campaigns (Leskovec, Adamic, & Huberman, 2007). Techniques and metrics related to the analysis of network dynamics and information propagation are highly active areas of research, particularly in technology-mediated networks (Kleinberg, 2008).

Early techniques that examine the spread of disease through social networks have been extended to better understand the spread of other phenomena such as happiness (Fowler & Christakis, 2008), obesity (Christakis & Fowler, 2007), information (Haythornthwaite, 1996), and innovations (Rogers, 1995). Increasingly, social media systems such as Twitter and Facebook are used to facilitate the flow of information, allowing researchers to examine information diffusion at a scale never before possible (Bakshy et al., 2012; Kwak, Lee, Park, & Moon, 2010). These observations serve as the foundation of theoretical models that explain information

dissemination (see Kleinberg, 2008, for an introduction and additional resources). As "Google + Ripples" and comparable Twitter visualization tools become available to more closely track the flow of content through CSCW systems, practitioners will be able to better understand how certain ideas spread and perform tests to identify what leads to increased spread of information.

In addition to information flowing through networks, network structures themselves can change: people make new friends or break up with old ones, employees get hired and fired, and users change who they communicate with. For example, findings inferred from e-mail exchanges suggest that existing network topology and organizational structures shape changes in social networks (Kossinets & Watts, 2006). Researchers examine changes in networks in many ways ranging from comparing network metrics from different snapshots in time to highlighting important critical events or "bursts" in the network (Barabasi, 2010) and using computational models to simulate network changes over time. Additionally, specialized network visualization tools allow researchers to examine changes to networks over time (Ahn, Taieb-Maimon, Sopan, Plaisant, & Shneiderman, 2011). Dynamic analysis features are increasingly being added to existing network tools as well, which often allow edges and vertices to be timestamped so that network growth can be "played back."

Network Visualization

Social networks are often best understood through visualizations, which can provide both insights and inspiration. As Fig. 1 shows, visual properties such as color, size, and positioning of the nodes can highlight important nodes, subgroups, and overall network properties. However, creating meaningful network visualizations is not trivial. It involves an iterative process of filtering out nodes and edges, mapping network metrics to appropriate visual properties such as size and color, laying out the nodes in a way that shows inherent structure and network motifs (e.g., via force-directed layouts), and labeling important nodes and edges (Hansen et al., 2009).

Ideally, networks will attain "netviz nirvana" (Bonsignore et al., 2009) wherein the following goals are achieved:

- Every vertex is visible.
- Every vertex's degree is countable.
- Every edge can be followed from source to destination.
- Clusters and outliers are identifiable.
- Unnecessary edge crossings are removed.

Tools like Gephi and NodeXL provide a range of features and built-in layout algorithms that help reach these goals for most networks with vertices in the hundreds or low thousands, though larger and/or denser networks pose significant challenges. Current research is exploring the use of network readability metrics (Dunne & Shneiderman, 2009), techniques that combine network visualization and

statistical overlays (Perer & Shneiderman, 2008), and graph summarization techniques (Dunne & Shneiderman, 2012) that may help us gain insights from visualizations of much larger networks in the future.

What Constitutes Good Work

Because SNA is performed by so many different communities of practice, there is a range of different expectations and criteria for determining what constitutes acceptable work. As SNA becomes more widespread in the HCI community, it is important to report appropriate metrics and use valid statistical techniques to validate claims, as opposed to simply presenting network visualizations. Below are a few best practices that apply to most SNA projects:

- Use network metrics that are appropriate for the type of network being examined. For example, if you are analyzing a directed network then in-degree and out-degree should be reported as opposed to degree. Likewise, if the network is weighted then, when possible, versions of network metrics that take the weights into consideration should be used. Where this is not possible, authors should state the reasons for using the basic, unweighted metric and associated limitations and implications.
- Do not claim more than your data can support. Network data, particularly collected from CSCW systems, is necessarily a simplification of much more complex social relations. Do not assume that Facebook friendships or e-mail exchanges necessarily equate to real-world friendships or that Twitter users are representative of the US population.
- Customize your network visualizations to illustrate the core points you are making (see Network Visualization section above for details). Remember that different network layout algorithms will highlight different properties of a network, so network visualizations should be used in conjunction with network metrics and statistical techniques.
- Use appropriate statistical techniques when mapping network properties to outcomes of interest or comparing networks. Though beyond the scope of this chapter, it is important to recognize that unique statistical techniques must be used when working with network data. For example, networks are often compared to a baseline network model (of which there are many) to demonstrate that certain features occur more often than expected. See Butts (2008) for a nice overview and introduction.
- Look at exemplary work, such as the articles cited throughout this chapter, for examples of methods and techniques appropriate for your questions. High-quality HCI work is often found in the CSCW conference, ICWSM conference, and CHI, while SNA articles using recent methods are found in *Social Networks: An International Journal of Structural Analysis*, the *Journal of Social Structure (JOSS)*, and *Connections*.

Additional Resources

The following annotated bibliography includes some good resources for becoming more expert in SNA. Books that require no relevant background are listed first, progressing to books that are written and used by experts in the field:

- John S. (2000). *Social network analysis: A handbook* (2nd ed.). Sage Publications Ltd: This is an excellent starting point for understanding SNA theory and methods, which assumes no prior knowledge. Written from a sociology perspective.
- Hansen, D., Shneiderman, B., & Smith, M. (2011). *Analyzing social media networks with NodeXL: Insights from a connected world*. Morgan Kaufmann: This introductory text focuses on analyzing social media datasets and assumes no knowledge of network analysis. It includes a tutorial-style section that shows how to conduct network analysis using the NodeXL software package as well as case studies from leading researchers in the field. Written from an HCI and marketing perspective.
- Nooy, W. D., Mrvar, A., & Batagelj, V. *Exploratory social network analysis with Pajek (Structural analysis in the social sciences)*. Revised and expanded second edition. Cambridge University Press: This introductory text introduces readers to a range of analysis techniques that can be performed by the Pajek software. Example datasets and exercises accompany the text as well as an appendix that walks the readers through the use of Pajek itself. Written from a mathematical and sociology perspective.
- Newman, M. (2010). *Networks: An introduction*. Oxford University Press: This comprehensive reference-style textbook introduces readers to the mathematics, theory, and algorithms used to analyze, model, and describe networks. Written from a physics and computer science perspective.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications (Structural analysis in the social sciences)*. Cambridge University Press: Long considered the "bible" of SNA, this book is a comprehensive mathematically focused reference book on SNA techniques. Written from a mathematical and sociology perspective.
- Scott, J. P., & Carrington, P. J. (2011) *The SAGE handbook of social network analysis*: This reference book includes chapters on major SNA topics (e.g., social support, cyber communities, terrorist networks) and methods (network surveys, sampling, statistical models, dynamic network analysis) written by leading authors in the field.
- Newman, M., Barabási, A., & Watts, D. J. (2006). *The structure and dynamics of networks*: This edited volume covers recent developments in SNA from leading authors in the field. Articles cover historical developments, empirical studies, modeling networks, and various application domains.

How the Authors Became Enamored with Social Network Analysis

Derek Hansen

My introduction to network theory began while browsing the shelves at the original Borders in Ann Arbor as a graduate student. I came across Duncan Watt's book "Six Degrees: The Science of a Connected Age" and read half of it that night in the store. I immediately recognized its potential for understanding interactions occurring in online communities, the focus of my research. As I began teaching at the University of Maryland's iSchool I started collaborating with Ben Shneiderman and Marc Smith on the evaluation and development of the newly created NodeXL network analysis tool. As HCI researchers, we saw our role as "democratizing" network analysis by developing a tool that would help make SNA accessible to a much wider audience. Not only could researchers use NodeXL to answer compelling social science questions, but also practitioners such as online community managers could use it to gain actionable insights into their own communities. I have been amazed at how quickly my students can now adopt "network thinking" and develop compelling network visualizations and analyses that tell important stories. No longer is the analysis of relational data relegated to a backstage role because of its obscurity. It can now take its rightful position on center stage alongside other methods that analyze more traditional qualitative and quantitative data sources. I see a bright future ahead for SNA, particularly as it is integrated with other methods as when researchers use SNA to identify the people they should interview or salient topics discussed by users within a similar network cluster. As an HCI researcher, I am particularly anxious to see collaborative system designers apply SNA as a tool to evaluate, understand, and design better systems.

Marc Smith

I have been interested in social uses of technology for many years, starting with bulletin-board systems accessed with dial-up modems. As a sociologist, I want to understand social media and be able to visualize the complex relationships, structures, and changes that are possible there. I use network analysis with a range of visualization techniques to create insights into the shape and structure of social media. I think of it as a kind of hashtag or keyword group photo. I take many pictures of many groups, and I look for patterns in the network as a whole, its subgroups, and the key people within those groups. I compare many networks of the same topic or compare topics to one another. I find that there are many different types of networks in social media and that there are different roles within those networks that are occupied by key people in strategic locations. I can now tell some stories about the size and shape and key people and subgroups within social media

topics. For example, political discussions in the United States are highly polarized with highly dense but separated groups, but this pattern is less visible in other nations' political discussions. Commercial discussions are often distinct from political topics; conversations about brands are often sparse even when they attract a large population. People mentioning these brands often have no connection to one another. In contrast, some products have formed communities, with populations with dense interconnections. Within communities there are often few people occupying the position of hubs and bridges. People at the center speak more often and have many more connections. Bridges often have fewer connections than hubs but have connections that reach from their own cluster across to many other clusters.

Exercises

Many kinds of things and relationships can be represented by nodes and links in SNA. Describe a network embedded within a CSCW system (e.g., Facebook's wall post network; Twitter's Follow network, Instagram's "Like" network) by describing what the nodes and edges mean, as well as the type of network (directed/undirected; weighted/unweighted; uniplex/multiplex; unimodal/multimodal;partial/complete).

Based on the network chosen above, describe actionable insights that could be gained from:

- Calculating a node-specific metric (e.g., Betweenness Centrality),
- Calculating an aggregate network metric (e.g., Density),
- Identifying clusters (e.g., subgroups) in the network, and/or
- Measuring network dynamics and/or information flow in the network.

References

Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 U.S. election: divided they blog. *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD'05 (pp. 36–43). New York, NY: ACM.

Ahn, J., Taieb-Maimon, M., Sopan, A., Plaisant, C., & Shneiderman, B. (2011). Temporal visualization of social network dynamics: Prototypes for nation of neighbors. In J. Salerno, S. Yang, D. Nau, & S.-K. Chai (Eds.), *Social computing, behavioral-cultural modeling and prediction* (Lecture notes in computer science, Vol. 6589, pp. 309–316). Berlin: Springer.

Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. *Proceedings of the 21st International Conference on World Wide Web*, WWW'12 (pp. 519–528). New York, NY: ACM.

Barabasi, A. L. (2010). *Bursts: The hidden pattern behind everything we do* (1st ed.). New York: Dutton Adult.

Beyer, H., & Holtzblatt, K. (1997). *Contextual design: Defining customer-centered systems* (1st ed.). San Francisco, CA: Morgan Kaufmann.

Bonsignore, E. M., Dunne, C., Rotman, D., Smith, M., Capone, T., Hansen, D. L. et al. (2009). First steps to Netviz Nirvana: Evaluating social network analysis with NodeXL. *International*

Conference on Computational Science and Engineering, 2009. CSE'09. (Vol. 4, pp. 332–339).

Borgatti, S. P., & Foster, P. C. (2003). The network paradigm in organizational research: A review and typology. *Journal of Management, 29*(6), 991–1013.

Burt, R. (1995). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.

Butts, C. T. (2008). Social network analysis: A methodological introduction. *Asian Journal of Social Psychology, 11*(1), 13–41.

Chau, D. H., Kittur, A., Hong, J. I., & Faloutsos, C. (2011). Apolo: Making sense of large network data by combining rich user interaction and machine learning. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, CHI'11 (pp. 167–176). New York, NY: ACM.

Chen, J., Geyer, W., Dugan, C., Muller, M., & Guy, I. (2009). Make new friends, but keep the old: Recommending people on social networking sites. *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, CHI'09 (pp. 201–210). New York, NY: ACM.

Christakis, N. A., & Fowler, J. H. (2007). The spread of obesity in a large social network over 32 years. *New England Journal of Medicine, 357*(4), 370–379.

Cross, R., Parker, A., & Borgatti, S. P. (2002). A bird's-eye view: Using social network analysis to improve knowledge creation and sharing. *IBM Institute for Business Value*. Retrieved Jan 1, 2006, from http://www1.ibm.com/services/us/imc/pdf/g510166900abirdseyeviewusing socialnetworkanalysis.pdf.

Cross, R., Parker, A., Prusak, L., & Borgatti, S. (2001). Knowing what we know: Supporting knowledge creation and sharing in social networks. *Organizational Dynamics, 30*(2), 100–120.

Dawson, S. (2010). "Seeing" the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology, 41*(5), 736–752.

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2006). "Alone together?": Exploring the social dynamics of massively multiplayer online games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI'06 (pp. 407–416). New York, NY: ACM.

Ducheneaut, N., Yee, N., Nickell, E., & Moore, R. J. (2007). The life and death of online gaming communities: a look at guilds in world of warcraft. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI'07 (pp. 839–848). New York, NY: ACM.

Dunne, C., & Shneiderman, B. (2009). Improving graph drawing readability by incorporating readability metrics: A software tool for network analysts. *University of Maryland, HCIL Tech Report HCIL-2009–13*. Retrieved April 1, 2014, from http://www-lb.cs.umd.edu/~cdunne/hcil/pubs/Dunne09Improvinggraphdrawing.pdf.

Dunne, C., & Shneiderman, B. (2012). Motif simplification: Improving network visualization readability with fan and parallel glyphs. *University of Maryland, HCIL Tech Report HCIL-2012-11*. Retrieved April 1, 2014, from http://www.cs.umd.edu/~cdunne/hcil/pubs/Dunne12Motifsimplification_Improving.pdf.

Eagle, N., Pentland, A. (. S.)., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. *Proceedings of the National Academy of Sciences, 106*(36), 15274–15278.

Ehrlich, K., Lin, C.-Y., & Griffiths-Fisher, V. (2007). Searching for experts in the enterprise: Combining text and social network analysis. *Proceedings of the 2007 International ACM Conference on Supporting Group Work*, GROUP' 07 (pp. 117–126). New York, NY: ACM.

Eveland, J. D., Blanchard, A., Brown, W., & Mattocks, J. (1994). The role of "help networks" in facilitating use of CSCW tools. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, CSCW'94 (pp. 265–274). New York, NY: ACM.

Farooq, U., Ganoe, C. H., Carroll, J. M., & Giles, C. L. (2007). Supporting distributed scientific collaboration: Implications for designing the CiteSeer collaboratory. *System Sciences, 2007. HICSS 2007. 40th Annual Hawaii International Conference on* (pp. 26–26). http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=4076423.

Fisher, D., & Dourish, P. (2004). Social and temporal structures in everyday collaboration. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI'04 (pp. 551–558). New York, NY: ACM.

Fowler, J. H., & Christakis, N. A. (2008). Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *British Medical Journal, 337*, a 2338.

Freeman, L. C. (2004). *The development of social network analysis: A study in the sociology of science*. Vancouver: Empirical Press.

Gilbert, E. (2012a). Predicting tie strength in a new medium. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work*, CSCW'12 (pp. 1047–1056). New York, NY: ACM.

Gilbert, E. (2012b). Designing social translucence over social networks. *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems*, CHI'12 (pp. 2731–2740). New York, NY: ACM.

Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, CHI'09 (pp. 211–220). New York, NY: ACM.

Golbeck, J., & Hansen, D. (2011). Computing political preference among twitter followers. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, CHI'11 (pp. 1105–1108). New York, NY: ACM.

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1360–1380.

Hanneman, R. A., & Riddle, M. (2005). *Introduction to social network methods*. University of California Riverside. Retrieved April 1, 2014, from http://faculty.ucr.edu/~hanneman/nettext/.

Hansen, D. L., & Johnson, C. (2012). Veiled viral marketing: disseminating information on stigmatized illnesses via social networking sites. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI'12 (pp. 247–254). New York, NY: ACM.

Hansen, D. L., Rotman, D., Bonsignore, E., Milic-Frayling, N., Rodrigues, E. M., Smith, M., et al. (2009). Do you know the way to SNA?: A process model for analyzing and visualizing social media data. *Human Computer Interaction Lab Tech Report HCIL-2009–17. University of Maryland*.

Hansen, D., Shneiderman, B., & Smith, M. A. (2010). *Analyzing social media networks with NodeXL: Insights from a connected world* (1st ed.). San Francisco, CA: Morgan Kaufmann.

Haythornthwaite, C. (1996). Social network analysis: An approach and technique for the study of information exchange. *Library & Information Science Research, 18*(4), 323–342.

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

Himelboim, I., Hansen, D., & Bowser, A. (2012). Playing in the same Twitter network. *Information, Communication & Society*, 1–24. http://www.tandfonline.com/doi/abs/10.1080/13691 18X.2012.706316#.UzpGVa1dUSQ.

Kairam, S. R., Wang, D. J., & Leskovec, J. (2012). The life and death of online groups: Predicting group growth and longevity. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM'12 (pp. 673–682). New York, NY: ACM.

Keegan, B., Gergle, D., & Contractor, N. (2012). Staying in the loop: Structure and dynamics of Wikipedia's breaking news collaborations. http://dl.acm.org/citation.cfm?id=2462934.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'03 (pp. 137–146). New York, NY: ACM.

Kivran-Swaine, F., Govindan, P., & Naaman, M. (2011). The impact of network structure on breaking ties in online social networks: unfollowing on twitter. *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems*, CHI'11 (pp. 1101–1104). New York, NY: ACM.

Kleinberg, J. (2008). The convergence of social and technological networks. *Communications of the ACM, 51*(11), 66–72.

Kossinets, G., & Watts, D. J. (2006). Empirical analysis of an evolving social network. *Science, 311*(5757), 88–90.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web*, WWW'10 (pp. 591–600). New York, NY: ACM.

Leskovec, J., Adamic, L. A., & Huberman, B. A. (2007). The dynamics of viral marketing. *ACM Transactions on the Web, 1*, 1.

Leskovec, J., & Faloutsos, C. (2006). Sampling from large graphs. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD'06 (pp. 631–636). New York, NY: ACM.

Leskovec, J., & Horvitz, E. (2008). Planetary-scale views on a large instant-messaging network. *Proceedings of the 17th International Conference on World Wide Web*, WWW'08 (pp. 915–924). New York, NY: ACM.

Marsden, P. V. (2005). Recent developments in network measurement. In P. J. Carrington, J. Scott, & S. Wasserman (Eds.), *Models and methods in social network analysis* (Structural analysis in the social sciences). New York, NY: Cambridge University Press.

McDonald, D. W. (2003). Recommending collaboration with social networks: A comparative evaluation. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI'03 (pp. 593–600). New York, NY: ACM.

Milgram, S. (1967). The small world problem. *Psychology Today, 2*, 60–67.

Müller-Prothmann, T. (2006). Leveraging knowledge communication for innovation. *Framework, Methods and Applications of Social Network Analysis in Research and Development.* Frankfurt a. M. et al.: Peter Lang.

Munson, S. A., & Resnick, P. (2010). Presenting diverse political opinions: How and how much. *Proceedings of the 28th International Conference on Human Factors in Computing Systems*, CHI'10 (pp. 1457–1466). New York, NY: ACM.

Newman, M. (2010). *Networks: An introduction* (1st ed.). New York, NY: Oxford University Press.

Newman, M., Barabási, A. L., & Watts, D. J. (2006). *The structure and dynamics of networks* (1st ed.). Princeton, NJ: Princeton University Press.

Perer, A., & Guy, I. (2012). SaNDVis: Visual social network analytics for the enterprise. *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion*, CSCW'12 (pp. 275–276). New York, NY: ACM.

Perer, A., & Shneiderman, B. (2008). Integrating statistics and visualization: Case studies of gaining clarity during exploratory data analysis. *Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems*, CHI'08 (pp. 265–274). New York, NY: ACM.

Pickering, J. M., & King, J. L. (1992). Hardwiring weak ties: Individual and institutional issues in computer mediated communication. *Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work*, CSCW'92 (pp. 356–361). New York, NY: ACM.

Rogers, E. M. (1995). *Diffusion of innovations*. New York: Simon and Schuster.

Steinfield, C., DiMicco, J. M., Ellison, N. B., & Lampe, C. (2009). Bowling online: Social networking and social capital within the organization. *Proceedings of the Fourth International Conference on Communities and Technologies*, C&T'09 (pp. 245–254). New York, NY: ACM.

Ugander, J., Karrer, B., Backstrom, L., & Marlow, C. (2011). The anatomy of the Facebook social graph. *arXiv:1111.4503*. Retrieved April 1, 2014, from http://arxiv.org/abs/1111.4503.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications* (1st ed.). Cambridge: Cambridge University Press.

Wellman, B. (2001). Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research, 25*(2), 227–252.

Welser, H. T., Gleave, E., Fisher, D., & Smith, M. (2007). Visualizing the signatures of social roles in online discussion groups. *Journal of Social Structure, 8*(2), 1–32.

Welser, H. T., Underwood, P., Cosley, D., Hansen, D., & Black, L. (2010). Wiki networks: connections of creativity and collaboration. In D. L. Hansen, B. Shneiderman, & M. Smith (Eds.), *Analyzing social media networks with NodeXL* (pp. 247–271). New York: Morgan Kaufmann.

Zaphiris, P., & Sarwar, R. (2006). Trends, similarities, and differences in the usage of teen and senior public online newsgroups. *ACM Transactions on Computer-Human Interaction, 13*(3), 403–422.

Zhang, J., Ackerman, M. S., & Adamic, L. (2007). Expertise networks in online communities: Structure and algorithms. *Proceedings of the 16th International Conference on World Wide Web*, WWW'07 (pp. 221–230). New York, NY: ACM.