Analysis of Social Networks & Group Dynamics from Electronic Communication

Nishith Pathak, Sandeep Mane, Jaideep Srivastava
Department of Computer Science,
University of Minnesota, Minneapolis, USA
{npathak, smane, srivasta}@cs.umn.edu

1. Introduction

The field of social network analysis evolved from the need to understand social relationship and interactions within a group of individuals. Knowing all individuals (employees) in an organization is difficult for an employee due to his/her limited bandwidth. Thus, in an organization's social network, not everyone directly knows (or interacts) with each other (Cross et al., 2002). Nor does an individual observe all the communication between individuals known (or unknown) to him/her directly. The result is that each individual forms perceptions about communication between other individuals, and uses them in his/her daily tasks. Having correct perceptions for all individuals in the organization is of utmost importance for the proper functioning of business processes.

Cognitive analysis of social networks has grown out of this interest in understanding what an individual's perceptions are about other individuals in terms of who they know (*socio-cognitive network analysis*), or what knowledge they have (*cognitive-knowledge network analysis*) (Wasserman and Faust, 1994). Traditional cognitive analysis approaches depend on the use of surveys and feedback from individuals. However, the lack of inability to collect large datasets, as well as problems such as inherent bias in responses, makes it difficult to analyze such social networks on a large scale. The widespread adoption of computer networks in organizations and the use of electronic communication for business processes have fostered a new age in social network analysis. E-mail communication, for example, is widely used by employees to exchange information. An email server observes all such communication between individuals in the organization, and therefore can analyze the email logs to determine the perceived social network for each individual, as well as the gold standard (or ground truth or real) social network. Given the large dataset sizes, it is difficult to apply existing techniques, since they do not scale very well. Hence, new efficient, scalable techniques are required for the socio-cognitive network analysis.

First, the problem of socio-cognitive analysis of a social network is presented. This is described using email communication network, and then our previous simple yet scalable approach is presented for such analysis. The approach can likewise be applied to other communications like instant messages. Previous case study using the proposed approaches on Enron email logs is then described. It uses the Enron email dataset, wherein the email communication between the employees of Enron is analyzed using the email logs before and after the Enron crisis of 2001.

The second part of the paper describes the problem of modeling and analysis of group dynamics in a social network. Data logs from a multi-player network based game, Sony EverQuest2, are now available, and are part of our current research on group dynamics. A brief overview of this problem is described and current research directions are explained.

2. Cognitive Networks & Their Analysis

(a) What are Cognitive Networks?

A *social network* is informally defined as a group of actors (or individuals) interacting with each other. Actors' cognitions make up an important component of any social environment and have a profound impact on the social network structure as well as actions of actors in this

network. *Cognitive networks* refer to the representation of actors' social interactions; as well as their perceptions which affect and/or are affected by such social interactions. The notion of "perceptions affecting social interactions" is used very broadly and can represent multiple things. For example, if one is studying a group of people in a book reading club, then cognitions pertaining to their social links can be a sub-set of each actor's characteristics, such as preferred genre, preferred authors, objective of joining the book club (e.g., casual hobby, or research). For each actor, these factors affect/are affected by his/her social links. The definition of these perceptions depends on the social network as well as the objective for such analysis. Research involving cognitive networks examines the complex relationship between actors' cognition and social structure. Study of cognitive networks is generally specific to the underlying social network. However, there is a special case of cognitive networks that is of particular interest. It deals with what each actor, in a social network, perceives about the social influence/connections of other actors in the network. Such a network is called a *socio-cognitive network* (or informally as a 'who knows who knows who' network) (Contractor, 2006) and is the main focus of our work.

(b) Significance of Cognitive Networks

Cognitive networks are an important component of group behavior analysis and have been studied in the fields of social-psychology, organizational dynamics and consumer research (Kilduff and Tsai, 2003)**,** (Ward and Reingen, 1990)**.** These are particularly used for the analysis of group consensus development or group decision making (Kameda et al, 1997), which profoundly impacts the development of techniques for predicting group behavior as well as analyzing the stability of a social group. The focus of our work is in the domain of organizational dynamics. Individual perceptions regarding other actors' social influence plays a very important part in determining the 'health' of any group. For example, it helps to identify individuals who command various factions within an organization, thus indirectly identifying/mapping the perceived 'informal hierarchy' in the network. These informal roles assigned arise due to the perceptions of fellow actors in the group (Krackhardt, 1993), and is a useful piece of information to a manager, as it allows him/her in building consensus regarding roles of individuals within the group, thus enabling its smooth functioning. For example, it may be necessary to build a consensus within a group regarding the leadership skills of an actor if some other less suitable actor is perceived as a better leader, or if possible, to promote a system where the better leader acts as mentor for the actor who commands respect among individuals thus, leveraging both actors' capabilities. Thus, the research will be useful in identifying upcoming leaders in a group, prediction of people changing factions, and so on.

3. Socio-cognitive Network Analysis: Techniques and Applications

This section describes our socio-cognitive network analysis techniques applied to the Enron email corpus (Pathak et al, 2006).

(a) Socio-cognitive Network Analysis

For socio-cognitive network analysis, the first phase is the modeling phase where the social network is modeled as a simplistic model that assumes independence between individual actors' observations. This model enables one to examine certain basic issues such as data sparseness and dynamic nature of the cognitive states as well as allows one to define basic distance measures for cognitive states. It also serves as a baseline model, which can be compared to more enhanced, complicated models. In the model, an observed instance of communication is treated as a Bernoulli trial and every actor estimates a Bernoulli parameter for an actor communicating with another actor. Since the Bernoulli parameters are dynamic, a time windowing scheme is used to address non-stationarity. A Bayesian update scheme is used to combine the information gained in one time window with the history of observations. An actor updates his/her perceptions based only on the emails he/she

receives and sends. A special actor (e.g., email server) that simply observes every email exchanged in the social network is added and its perceptions are used "ground truth". In the analysis phase, a similarity measure is used to identify groups of actors who have similar perceptions and identifying actors close to the ground truth. In our study, KL-divergence based measures are used to define similarity between two actors' perceptions.

(b) Example 1: Enron Email Corpus

The socio-cognitive network in Enron is modeled using the Enron email corpus and results of the socio-cognitive network immediately after the Enron crisis is compared with the socio-cognitive network one year before the crisis.

A graph, called "*agreement graph*", is constructed where nodes represent actors and an edge exists between two nodes if the similarity between their perceptions is more than a user defined threshold (see Pathak et al, 2006a for details on experimental settings). Figure 1 shows the agreement graph for October, 2000. It is observed that the graph consists of many small, disjoint components of users. A reason for this is that big organizations like Enron usually have many organizational groups with high intra-group and low inter-group communication. Figure 2 shows the agreement graph for October 2001, which mainly consists of one large, connected component. This indicates that there is a considerable similarity in social perceptions during the crisis period. The connectivity of the October 2001 agreement graph also indicates that communication (and hence information) is shared among various actors. Such a network is highly conducive towards dissemination of ideas in a social network. In case of Enron, the crisis and related subjects were "hot topics", i.e., were often discussed in the underlying social network. Note that the number of nodes in the October 2001 graph is much more than that of the October 2000 graph. This is because an actor is included in the agreement graph only if his/her perceptions have similarity greater than the threshold with at least one other actor.

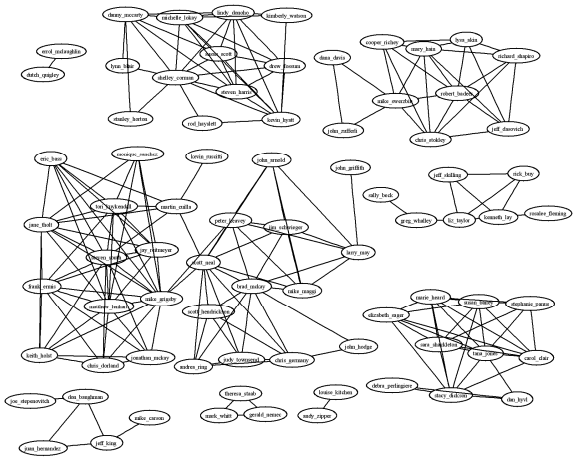

Figure 1. Agreement Graph for October 2000

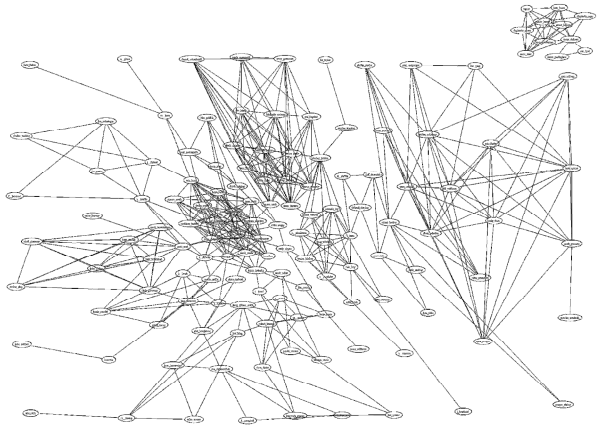

Figure 2. Agreement Graph for October 2001

To analyze the actors' perceptions closeness to reality, the similarity measure (termed "*r-closeness*") between the actors' perceptions and ground truth is computed, and then the actors are ranked in decreasing order of r-closeness. For October, 2000, the actors are roughly divided into three categories - actors who are communicatively active and observe a lot of diverse communications, (these actors occupy the top rankings), actors who also observe a lot of communication; however, their skewed observations lead to skewed perceptions (they occupy the medium range of ranks) and actors communicatively inactive and who hardly observe any of the communication (occupying the bottom ranks). Table 1 summarizes the percentages of various actors (categorized by their formal positions) in the different ranges of r-closeness rankings. Using the rankings for October 2000, a socio-cognitive network hypothesis of interest to sociologists is examined.

H1. Higher is an actor in the organizational hierarch, better is his/her perception of the social network.

From the r-closeness rankings, it is observed that majority of the top positions are not occupied by higher level executive employees. The top 50 ranks consist of a large chunk of the employee population (around 46.4% of the employees) along with 21.4% of the higher management and 34.4% of the executive management actors (see Table 1). These employees (most likely secretaries) are responsible for facilitating coordination and communication among multiple groups in the organization.

Table 1. Users in different rank ranges of r-closeness (October 2000)

Ranks	Not Avail-able	Employees	Higher Management	Executive Management	Others
1-10	10.3% (4)	4.9% (2)	7.1% (2)	3.4% (1)	7.1% (1)
11–50	17.9% (7)	41.5% (17)	14.3% (4)	31.0% (9)	21.4% (3)
51-151	71.8% (28)	53.6% (22)	78.6% (22)	65.6% (19)	71.5% (10)

Table 2 summarizes statistics for the r-closeness rankings for October 2001. These are significantly different from those of October 2000. The percentage of management staff among the top 50 ranks increased significantly at the cost of employees being pushed down. A possible reason for this is that during the crisis month, emails were exchanged across different levels of formal hierarchy in the organization thus exposing management level actors to more diverse communication (Deisner and Karley, 2005). Another intuitively appealing reason (Deisner and Karley, 2005) is that during October 2000, on an average, management people "sent" about 80% and "received" only 20% of the total communication they were exposed to. In the October 2001, there was a reversal and management people sent only 20% and received about 80% of their total communication. Thus, their perceptions improved as since they were observing (vs. initiating) a lot more communication during the later period. Finally, the volume of communication, for management level actors, in October 2001 was a lot more than in October 2000.

Table 2. Users in different rank ranges of r-closeness (October 2001)

Ranks	Not Avail -able	Employees	Higher Management	Executive Management	Others
1-10	5.1% (2)	2.5% (1)	3.6% (1)	20.7% (6)	0% (0)
11–50	23.1 % (9)	26.8% (11)	28.6% (8)	37.9% (11)	7.1% (1)
51-151	71.8% (28)	70.7% (29)	67.8% (19)	41.4% (12)	92.9% (13)

4. Group Dynamics: Motivation and Applications

This section describes the motivation to study the group dynamics in a large social network, and explains how this will be useful for analyzing the logs of Sony EverQuest2 network game.

(a) Group Dynamics

There is considerable evidence of the importance of groups that interact effectively for the global economy. The most important and complex decisions made by governments and organizations occur in group contexts. Consider the coverage of the group deliberations prior to the Bay of Pigs invasion or, more recently, the Challenger and Columbia shuttle tragedies. A central challenge, motivated by emerging developments in IT, is that the nature of groups and how they operate has changed radically. Spurred by globalization, virtual teams have become a staple in governments, businesses, and universities. Even before the recent IT revolution,

Hollywood and the construction engineering industry provided early historical examples of the emergence of *ad hoc* groups which brought together people with different skills from their latent networks for a specific task over a finite time period. Today an increasing preponderance of groups – in social, political, and economic contexts - are *ad hoc*, agile, flexible, transient entities that emerge from a larger primordial latent network of relationships for a short duration to accomplish a wide variety of tasks and then dissolve only to be reconstituted with a somewhat different configuration at same later point in time. While there is growing awareness of the socio-economic consequences of these groups, our understanding of the dynamics of the formation of these groups and their impact on group effectiveness is severely limited.

(b) Example 2: Sony EverQuest2 Logs

EverQuest 2 (EQ2) is one of the world's largest Massively Multiplayer Online (MMO) games, developed by Sony Online Entertainment. MMOs comprise tens of thousands of players who are at any one point in time coalescing in thousands of groups to accomplish "quests" and "raids" that involve a wide variety of activities similar to tasks we undertake in real life – finding information or materials; making, selling or buying products and services – and some that are less common such as "killing" other characters or plundering (http://everquest2.station.sony.com). Beyond the data collection challenge, the scale of the proposed research enterprise also poses significant computational challenges in uncovering and analyzing the complexities that govern the dynamics of group behavior in these virtual worlds.

The analysis seeks to address this limitation by developing a theoretical framework that extends extant research on group behavior to reflect the conceptualizations of groups in contemporary society. It proposes a network approach to modeling the eco-system of overlapping and constantly changing groups that constitute the fabric of contemporary society. It recognizes that empirically testing such a model poses formidable data collection challenges that have perhaps deterred scholars from embarking on such a theoretical excursion. However, a unique resource available to the research team provides access to all behavioral traces (server logs) from a virtual world particularly well-suited for theorizing and empirically modeling the dynamics of group behavior. This research aims at applying data mining techniques to analyze the logs of the game, EQ2 to better understand group formation and dynamics. Specific research questions to be addressed include:

- How do networks within the ecosystem of groups enable and constrain the formation of groups?
- How do micro-group processes influence group effectiveness and social identity?

This research promises to usher in a new generation of theorizing and research on the emergence and performance of groups in complex social settings that are unfettered by the constraints on the type, quality, and quantity of data we have been heretofore confined to use. It will expand our knowledge of how groups form and operate in larger ecosystems of groups, individuals, and organizations. The analysis of logs generated from Virtual Worlds poses novel challenges from a computational perspective. This requires an interdisciplinary investigation that will result in new (i) information models for modeling the Virtual World, (ii) data structuring and algorithmic techniques for data access and analysis, and (iii) techniques for computational efficiency. The knowledge and tools developed in this research will allow researchers to understand more fully, and practitioners to cultivate more effectively, the emergence and performance of ad hoc groups in contemporary society. It will also provide other disciplines with new computational and statistical modeling methodologies and tools, which should have considerable positive implications for future research in other disciplinary areas.

5. Conclusion

Recording of electronic communication, i.e. email logs, web logs, etc., has instrumented human processes like never before. Analysis of these logs is creating an unprecedented

opportunity to understand sociological and psychological processes at a granularity, which just a few years ago were unimaginable. In this paper we've described two ongoing projects at the University of Minnesota that explore aspects of this emerging field. First involves analyzing emails from the Enron dataset to understand socio-cognitive networks, and develop quantitative measures for concepts such as *perceptual congruence* and *perceptual realism*. Second is analyzing logs from Sony's EQ2 MMO game to understand the formation and dynamics of group behavior, especially for *ad hoc* groups.

References

1. N. Contractor. Multi-theoretical multilevel MTML models to study the emergence of networks, NetSci conference, 2006.
2. R. Cross, N. Nohria and A. Parker, Six myths about informal networks and how to overcome them, MIT Sloan Management Review 43(3) (2002) 67–75.
3. Diesner, J., & Carley, K.M. (2005). Exploration of Communication Networks from the Enron Email Corpus. *Proceedings of Workshop on Link Analysis, Counterterrorism and Security, SIAM International Conference on Data Mining 2005*.
4. Kameda, T., Y. Ohtsubo, M. Takezawa. (1997), Centrality in socio-cognitive networks and social influence: An illustration in a group decision-making context. Journal of Personality and Social Psychology 73(2): 296-309.
5. Kilduff M and Tsai W (2003), Social networks and organizations, Sage, London.
6. David Krackhardt and Jeffrey R. Hanson. Informal Networks: The Company Behind the Chart. Harvard Business Review, 1993.
7. N.Pathak, S.Mane, J.Srivastava. Who thinks who knows who? Socio-cognitive analysis of an Email Network. IEEE International Conference on Data Mining, 2006.
8. Ward, James C. and Peter H. Reingen (1990), "Sociocognitive Analysis of Group Decision Making Among Consumers," Journal of Consumer Research, 17 (3), 245-63.
9. S. Wasserman and K. Faust, Social Network Analysis — Methods and Applications 11 (Cambridge University Press, 1994).