National Research Council Canada
Conseil national de recherches Canada

Institute for Information Technology
Institut de technologie de l'information

NRC · CNRC

Automated Social Network Analysis for Collaborative Work *

Korba, L., Song, R., Yee, G., Patrick, A.
September 2006

* published in the Proceedings of the Third International Conference on Cooperative Design, Visuallization and Engineering (CDVE 2006). Palma de Mallorca, Spain. September 17-20, 2006. NRC 48732.

Canada

Automated Social Network Analysis for Collaborative Work[1]

Larry Korba, Ronggong Song, George Yee, Andrew Patrick

Institute for Information Technology, National Research Council of Canada
Building M-50, Montreal Road, Ottawa, Ontario K1A 0R6
{Larry.Korba, Ronggong.Song,
George.Yee, Andrew.Patrick}@nrc-cnrc.gc.ca
http://www.iit-iti.nrc-cnrc.gc.ca

Abstract. Inter-networked computers enable virtual collaborative work. In the course of interacting with one another, individuals send and receive messages and files of various sorts. This may be done within specialized collaborative work environments, or by simply employing a combination of different communication tools and applications. In the course of doing their work, collaborators perform different actions that create and/or otherwise manipulate digital artifacts that are related to different aspects of their collaboration. Social network analysis is used to develop a fuller understanding of interactions between people. We describe a software prototype of a tool that automatically measures and analyzes aspects of collaboration developing visualizations of likely social interactions. In this paper we describe the system, some early results, and several different possible applications of the technology.

1 Introduction

"Computer-supported cooperative work" (CSCW) refers to cooperative work carried out by one or more individuals with computer and network support [1]. CSCW is often used to refer to the situation where people work together in dynamically-formed groups to accomplish particular tasks. Many of the different tasks may be accomplished using various computer applications. The nature of the applications may include file sharing of design documentation, white board applications for brainstorming, messaging systems, video conferencing to facilitate more tangible connection between collaborators and email for asynchronous communication.

These different types of network applications are used at different times depending on which of four modes of CSCW operation [1, 2] participants prefer to use: synchronous, distributed synchronous, asynchronous and distributed asynchronous. Many software applications have been developed specifically for CSCW, while other desktop applications are evolving to include collaboration functions [3].

In this paper we describe the implementation of a system intended to determine, in real time, the development of social networks during collaborative work. This Auto-

[1] National Research Council Paper Number 48732

mated Social Network Analysis Technology (ASNAT) is an implementation and extension of work we have introduced previously [4]. As its key advantage, it is not reliant on the use of any particular design or visualization application. The technology has many different possible uses, some of which we describe in a subsequent section.

In this paper, section 2 describes the problem we are addressing. Section 3 presents our approach, from our abstraction model, to system implementation details. We discuss some results, sketching possible applications of ASNAT in Section 4. Section 5 provides a discussion and conclusions.

2 Problem Statement

Social Network Analysis (SNA) [5] is a very powerful tool used to help form a better understanding of how people work together. SNA has been applied for many different purposes [6, 7, 8, 9]. However it has been found especially useful in understanding how groups of people interact via electronic communication [10, 11] such as usenet news and email. Social network analysts seek to describe, as fully as possible, networks of relationships between people, including relationships maintained by computer or network mediated interactions, or for the allocation of resources needed for work. SNA display relationships as graphs, with nodes representing individuals, and edges representing interactions types. The degree and type of interactions may be represented by the lengths, colors, and widths of nodes and edges. The information used in a social network analysis is almost always gathered manually. Manual collection involves observation and recording of activities, and/or the use of questionnaires, interviews and diaries. The reliability of the SNA results is often questioned because of the complex nature of this process. For instance, incorrect reporting (intentional or not), often occurs when participants record their own activities. Different interactions may not be remembered equally as well. Statistical data processing tools have been used to discover interaction patterns [9]. Analysts have used these patterns to interpret and to uncover what aspects were responsible for successful collaboration (for instance for design purposes) [10]. On a wider scale, SNA has been used to determine how organizations operate.

Our work involves the development of a system for automatically discovering and analyzing social networks inferred from the digital artifacts created when people interact with one another over computers and computer networks.

3 Our Approach

Figure 1 provides a schematic overview of the ASNAT approach. People within organizations communicate and collaborate with one another, while working to achieve their objectives. Computer networks form a common means for communication and interaction. In a computer-mediated environment, people often don't communicate in real space and time with one another. Instead, they work together using digital means. Networks facilitate remote interaction. In the course of performing various computer

and network interactions, digital artifacts are created, shared, modified, moved and destroyed among collaborators. These artifacts may be files, email messages, instant text messages, mouse or keyboard actions, network activities, etc. The artifacts are linked to the collaborative process. Our approach involves discovering artifacts on each computer platform created in the context of collaborations and determining the relationships between artifacts on all computers across a network. In other words, ASNAT performs a reverse engineering process on digital artifacts to detect patterns of human collaboration.

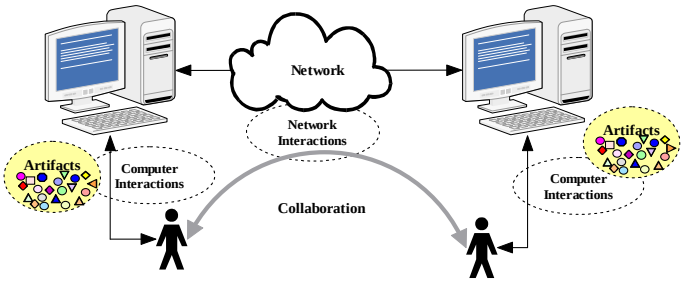

Figure 1. Human collaborations and computer activities

3.1 Design

In the context of ASNAT, a social network model is a schematic description of a set of social behaviors amongst users based upon digital artifacts developed within an enterprise. Figure 2. shows the ASNAT abstract design model.

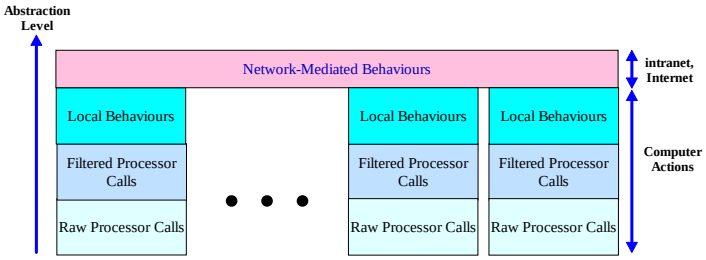

Figure 2. Abstraction model for ASNAT.

ASNAT uses a pair of software agents per computer platform. One agent monitors low level computer system calls, filtering them to produce a set of local artifacts related to local behaviors. The second agent accesses these artifacts, communicates with peers to determine the global relationships based upon a correlation with the artifacts of others. The correlation provides for fuzzy matching of artifacts.

ASNAT uses several components in the course of developing relationship diagrams. These include:

• A data structure to hold the filtered processor calls. There is a very large amount of raw data generated in a short period of time. To reduce the amount of data, unnecessary or duplicate information is filtered out retaining only activities that are collaboration inspired before any further processing.

• A set of rules/logic for creating and updating the relationship model. Based on the filtered data, ASNAT builds models representing local social behaviors. The local model provides views of social networks from an individual point of view.

• A set of methods (interfaces) for manipulating the model and interactions between relationship models. Based on local social behaviors in the third layer, ASNAT can build models representing social behaviors with an enterprise scope for real-time display.

• A graphical representation model for displaying the work flow and SNA models to a user. Currently there are several different software packages available for viewing and manipulating graphical representations of social networks, each with its own image format. We use an internal graphical representation that facilitates rapid display, manipulation, while offering easy conversion into other social network analysis formats (for instance Pajek [12])

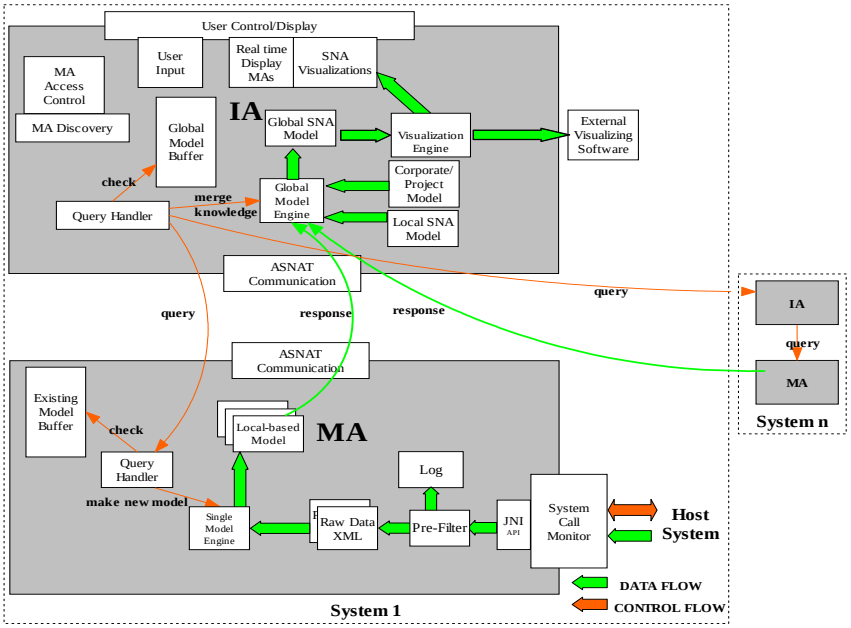

Figure 3. This diagram details the Monitoring Agent (MA) and the Interface Agent (IA) along with a simplified description of the information flows, shown for only one direction. In reality the IA handles bidirectional queries.

3.2 Implementation

ASNAT is implemented in C++, and Java [13] (on the Java Agent Development Environment (JADE) [14]). It is a decentralized design; using two agents resident on each computer. The Monitor Agent (MA) and the Interface Agent (IA) are illustrated schematically in Figure 3. The agents work together in the collection of artifact activity and for workflow and social network analysis. The Monitor Agent contains the sensing elements. A Java to Native language Interface (JNI) allows our Java applica-

tion to communicate with lower level system calls (written in C++). The system call monitor can monitor any computer activity (keyboard, mouse, window operations, network accesses, etc.). System call events are filtered to form the digital artifacts relevant for later analysis. The event information is analyzed to build a single user model, organizing artifacts that indicate possible social interactions. There may be many local models based upon the different types of artifacts being monitored.

The Interface Agent (IA) serves several purposes. One purpose is to build global models of networks-of-interest related to the local Monitoring Agent (MA). It does this by using the local models to formulate queries to be sent to other IAs through the IA's Query Handler. Responses from these queries are analyzed for fit within a global social network model based upon models for the local node. The Global Model Engine in turn handles queries from remote IAs to determine the correlation between local and remote artifacts The IA includes a visualization engine for workflow and social network models. A user interface provides many options for presenting and manipulating different social network analyses and visualizations. The IA also facilitates translation of ASNAT social networks into PAJEK format for other analyses.

4.0 Results and Applications

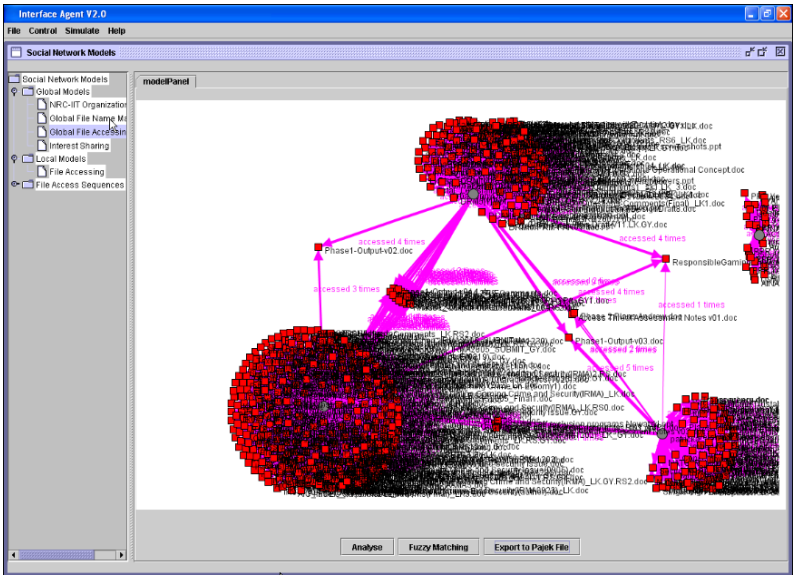

Figure 4. Screen snapshot of the user interface for the Interface Agent.

We have implemented many different sensors and filters for ASNAT. In this section we focus on artifacts related to files alone. Figure 4 is a screen snapshot of the user interface for the ASNAT Interface Agent. We use a multi-paned dialog box to select different views and to control the analysis of artifact relationships. The screen shot shows one graphical view of global file accesses by 4 different users (grey cir-

cles). The (red) squares indicate the different files a user has handled. The pink-colored edges with arrows between users point to red squares indicate files that were apparently shared between the users. In the left panel, an operator can chose different types of views.

Understanding workflow is important within a project or within an organization as a whole. ASNAT provides a visualization of workflow within an organization that can be superimposed upon an organizational structure as shown in Figure 5.

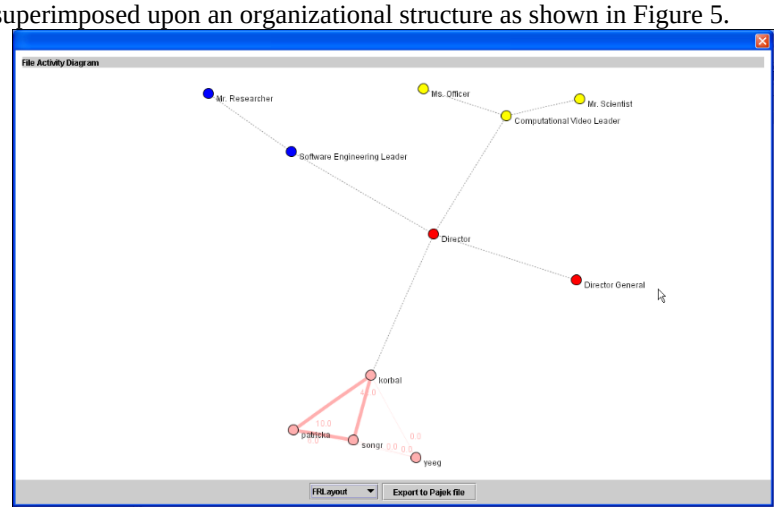

Figure 5. Screen snapshot from the Interface Agent showing organizational structure superimposed on the interactions between 4 different individuals.

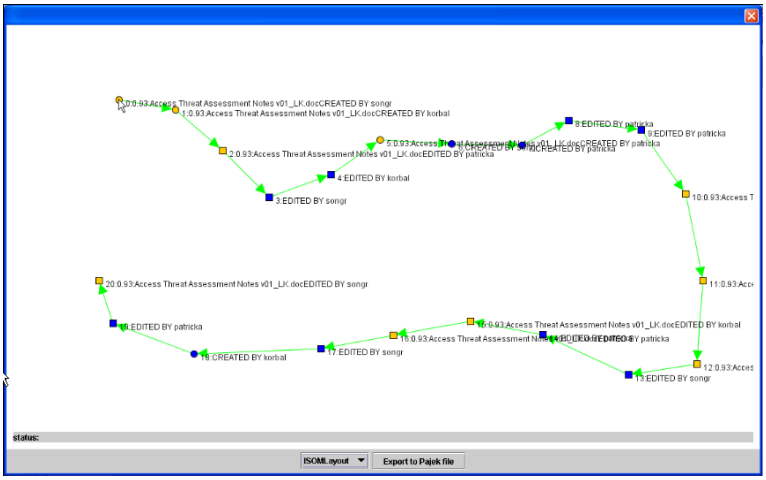

Figure 6. A screen snapshot for ASNAT showing activity across different users for one file over time. Contributions by different users are described at each node.

Another view implemented in our prototype is a graph of the activity related to one particular file artifact with the passage of time (see Figure 6). As individuals collaborate they generally create, edit, and share files. Sharing files may be accomplished using file transfer programs, network file servers, email, or other means. Clearly, individuals may also change the names of the files they receive and edit. ASNAT keeps track of these changes, logically linking files that may have had a similar origin. This is done both heuristically while users perform different file manipulation operations and by a fuzzy name matching technique in cases where ASNAT is determining relationships forged by examining file activity across different users' computers.

Automated workflow analysis using social networks as implemented in ASNAT has many possible applications. Due to lack of space we only briefly mention them here.

- By tracking interactions with files related to organizational operations (e.g. human resources), the system could map actual work flow amongst employees.
- ASNAT may also be used to assure that employees adhere to project or organizational policies, for instance for security, thus automating compliance.
- At the work-activity level, ASNAT could be linked with project management software to provide team members with a way of determining, at a glance, the status of different parts of the project, irrespective of the collaboration applications. This approach would also be useful for automating the discovery of expertise [15] as well as a tool for understanding the dynamic evolution of social relationships in groups [7]
- When equipped with text summarization or concept extracting software that extracts key contributions from different users in the documents they have produced, ASNAT could automatically produce or update knowledge profiles for organizational staff; an approach that would be helpful in determining expertise [8] or maintaining project/corporate memory.
- As illustrated in Figure 6, ASNAT could produce a type of forensic display associated with building a design-related document. This would be useful to determine what was done, by whom and when - invaluable information for all project members.
- With the addition of software for distributed replication of files produced by team members, ASNAT could become a distributed repository for many different versions of documents associated with a design, allowing users to easily access any version of a design at any time from anywhere on the corporate network.
- ASNAT could be the basis of a system to detect and subvert inappropriate insider behavior. Normal workflow for individuals in different organizational positions may be compared automatically against those for a person under suspicion in a similar position in an organization.

5 Conclusions

In this paper we have described the development of ASNAT, a system that embodies automated social network analysis technology. ASNAT has several possible applications. However, it is clear that in order to be successfully deployed, the functionality of the application must be carefully considered to assure that the system is not used

for unbridled monitoring of employee behavior. Currently we are extending the prototype for security and privacy compliance. This work also involves adding security mechanisms to prevent leakage of corporate data through ASNAT operations.

6 Acknowledgements

The authors acknowledge the Communications Security Establishment for its support in this work. We also acknowledge the assistance of Dr. Yuefei, Yungfeng Li, Tian Yuan, Hui Wang, Myroslav Palenychka, Ravi Chauhan, and Lun Howe Mark.

References

1. B. Bruegge and A. Houghton, Computer-Supported Cooperative Work: Discussion Summary, *Software Engineering: Education and Practice International Conference*, pp. 516-519, 1996.
2. K.L. Mills, Introduction to the Electronic Symposium on Computer-Supported Cooperative Work, A*CM Computing Surveys*, Vol. 31, No. 2, June, pp.105-115, 1999.
3. Ricadela, A., Office 2007 to feature new workflow and collaboration tools, Information Week, Feb. 15, 2006, available at: http://www.informationweek.com/shared/printableArticle.jhtml?articleID=180202580
4. Temdee, P., Korba, L. Of Networks, Interactions and Agents: An Approach for Social Network Analysis, *Proc. 6th International Conference on CSCW in Design*, London, Ontario, July 12-14, 2001
5. B. Wellman, For a social network analysis of computer networks: a Sociological Perspective on Collaborative Work and Virtual Community, *Proc. of the 1996 conference on ACM SIGCPR/SIGMIS Conference*, pp. 1-11, 1996.
6. Domingos, P., Richardson, M. Mining the Network Value of Customers, *Proc. of the 7th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Francisco, pp. 57-66, 2001.
7. Gloor, P. A. Capturing Team Dynamics through Temporal Social Surfaces, *Proc. of the 9th Int. Conf. on Information Visualization 2005*, July 6-8. 2005, pp. 939-944.
8. Zhang, J., Ackerman, M. Searching for Expertise in Social Networks: A Simulation of potential Strategies, *Proc. of the 2005 Int. ACM SIGGROUP conf. on Supporting Group Work*, Florida, 2005, pp. 71-80.
9. A. Popescul, L. Ungar. Statistical Relational Learning for Link Prediction, *Workshop on Learning Statistical Models from Relational Data*, IJCAI 2003.
10. Lin, F., Chen, C., Tsai, K. Discovering Group Interaction Patterns in a Teachers Professional Community, *Proc. 36th Ann. Hawaii Int. Conf. on System Science*, Jan. 6-9, 2003.
11. Fischer, D. Using Egocentric Networks to Understand Communication, *IEEE Internet Computing*, Vol 9, Iss. 5, Sept-Oct 2005, pp. 20-28.
12. Pajek: Program for Large Network Analysis Available at: http://vlado.fmf.uni-lj.si/pub/networks/pajek/
13. Java programming language. Available at: http://java.sun.com/
14. Jade Platform available at: http://sharon.cselt.it/projects/jade/ March, 2006.
15. Yu, B., Singh, M. Searching Social Networks, *Proc. of the 2nd Int. Joint Conf. on Autonomous Agents and Multiagent Systems*, Melbourne, Australia, 2003, pp. 65-72.