SAND2008-1298P

Feasibility Investigation of a Quantifiable and Objective Approach to Organizational Improvements at Sandia

A Seniors LDRD Proposal

by

Andrew J. Scholand, Ph.D.
05535 Data Analysis And Exploitation

Objectives

Sandia faces a tough future. Projected nuclear weapons funding is monotonically decreasing, and competition is severe in other areas of national security. Increasing efficiency and effectiveness is a critical priority if Sandia is to continue providing exceptional service in the national interest.

This proposal addresses technical means to make socially based efficiency improvements, based on the fundamental assumption that organizational and interpersonal improvements are likely to provide the highest gradient of returns at a technically skilled organization like Sandia. Furthermore, although much of Sandia's knowledge and expertise in successfully executing work quickly is not written down, informal conversations among and between groups of individuals are where this information is developed, shared, and acted upon in an operational context. Specifically, the fundamental proposition is that digital records arising from interacting is the 'as-is' organization can be analyzed to create an approximate but meaningful representation of the social dynamics within the organization. 'Meaningful' in this context implies facets of information relevant to interpersonal dynamics, aspects of distributed cognition and group work, and the development of organizational power and control. By constructing an explicit representation of working collectives of Sandians, it will be possible to better understand how work is actually accomplished, which in turn enables effective systemic improvements.[1]

An additional objective in proposing this work is to strengthen the strategic relationship Sandia has entered into with the University of Texas (UT) system through collaboration around this topic. Deepening the co-research relationship with UT is especially important moving forward because of the tremendous leverage of their information resources, particularly in areas Sandia has not been historically strong (e.g. the social sciences).

Technical Approach

This proposal moves toward this goal in the following manner. Work-related instant message conversations between up to 28 individuals in a SRN-hosted "public" chat room have been collected for more than a year. The proposed research will use directed conversations between individuals to construct an exemplar organizational construct, a sociogram. Sociograms, as shown in the Figure 1 example, are directed graphs of affective relations between individuals. Sociograms were developed by J.L. Moreno in the 1930s to analyze social relations from a perspective of the most fundamental unit of society being interpersonal ties rather than an individual. Solidarity, shared norms, identity, collective behavior, and social cohesion are considered to emerge from social relations. Understanding the structure of these relations has been shown to help explain

[1] For example, with an explicit representation of interaction patterns, management can organize teams so that relations between individuals within and across teams conform to suggestions from Social Balance Theory.

how information, attitudes and behavior diffuse within a system. In addition to this high-level view of the overall social structure, sociograms also provide insight into the role of individuals in the organization. Since individual-to-individual relations are necessary to have access to information and help, the number and intensity of a person's ties are called his or her sociability or social capital. Sociograms reveal those people who occupy central or strategic positions within the system of social relations and are crucial for the transmission process.

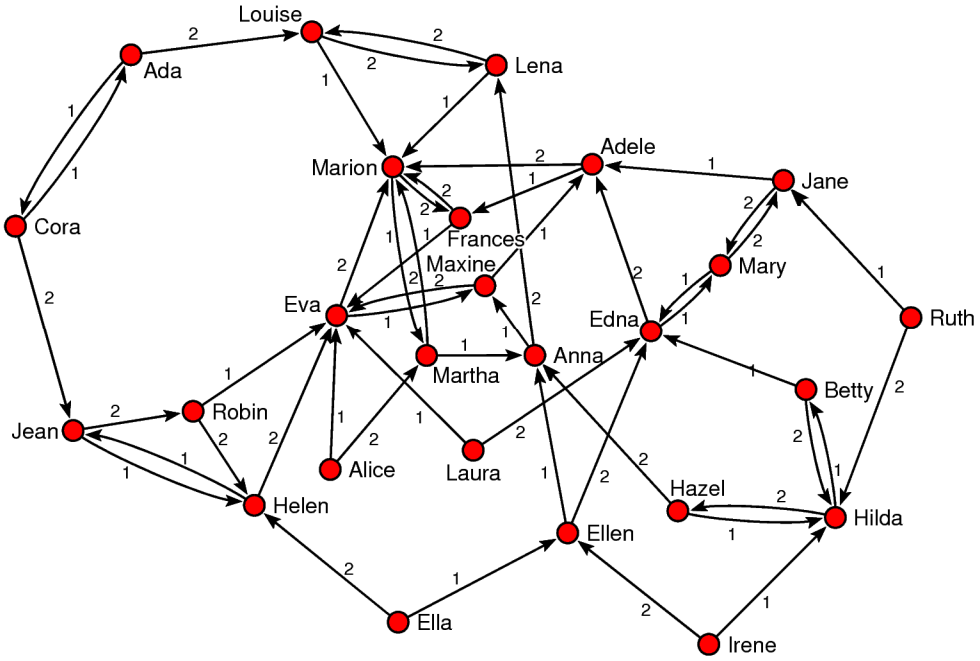

Figure 1: An example sociogram
(Source: W. de Nooy, A. Mrvar, V. Batagelj: *Exploratory Social Network Analysis with Pajek*, January 2005)

For purposes of comparison, we will use the traditional sociometric choice survey method to construct a sociogram among the participants. The key technical work, however, will be converting the time-stamped textual transcripts to graph form.

The initial working hypothesis will be that each person has a certain amount of daily discretionary attention that may be directed towards others. Patterns of interaction in the public forum (such as responses to queries and specific, directed verbal exchanges) can be considered to form pair-wise connections between members of the group. Using concepts of limited attention and the sustainability of strong and weak interpersonal ties, the relative strength of all possible pairings can be computed and compared. Linguistic analysis of the content of these links, moreover, can provide important insight into the richly layered and textured nature of each interpersonal working relationship.

The proposed work will leverage the Linguistic Inquiry and Word Count (LIWC) text analysis software program developed by University of Texas researchers James W. Pennebaker, Roger J. Booth, and Martha E. Francis. LIWC is a program for quantitative text analysis that uses a word count strategy for both the analysis of content (what is being said) and style (how it is being said). Word count strategies are based on the

assumption that the words people use convey psychological information over and above their literal meaning and independent of their semantic context. In this sense, they are "top down" in the sense that they explore text within the context of previously defined psychological content dimensions or word categories. (In contrast, word pattern strategies such as those used by the Sandia Text ANaLysis Extensible librarY – STANLEY -- mathematically detect "bottom-up" how words co-vary across large samples of text, typically to determine the degree to which two texts are similar in terms of their content.) LIWC searches for over 2300 words or word stems previously categorized by independent judges into over 70 linguistic dimensions. These dimensions include standard language categories (e.g., articles, prepositions, pronouns— including first person singular, first person plural, etc.), psychological processes (e.g., positive and negative emotion categories, cognitive processes such as use of causation words, self-discrepancies), relativity-related words (e.g., time, verb tense, motion, space), and traditional Freudian content dimensions. Researchers at the University of Texas would be funded to extend these categorizations to both lexical structures specific to instant messaging (e.g. phonetic abbreviations and 'emoticons') and to psychological markers of interpersonal dynamics between individuals in a work context, addressing differences arising from academic discipline, previous work history, and team experience.

Human Subject Board approval of aspects of the experiment design will be required for the proposed work, if funded. Preliminary discussions with Terry Reser have resulted in a placeholder for seeking such approval once funding is available, tracking number SNL0806.

Impact on State of Knowledge

Recent research (DiMicco and Yankelovich, 2007; Ehrlich et. al., 2007) has demonstrated substantial organizational value to social network informed information artifacts. To date, however, these approaches seek to leverage existing structures (co-authorship of papers, organizational and seating charts, email 'from' and 'to' headers) rather than build them from content analyzed with a particular theoretical viewpoint. Consequently, there have been no efforts to correlate automated social networks to their theoretical predecessors, sociograms. One outcome from this research will be a paper contrasting the traditional survey method and the proposed psycholinguistic analysis, to be submitted to the 2008 ACM conference on Computer Supported Cooperative Work (CSCW 2008).

Extending psychological word count analyses to the conventions and norms specific to instant messaging will extend the usefulness of the LIWC software distributed by the University of Texas.

Relevance to Sandia

Sandia stands to benefit in three primary ways from the proposed research. One, as described in the introduction, confirmation of the research hypothesis that highly accurate descriptions of the social dynamics of Sandians working on complex problems can be obtained by psycholinguistic analysis of the digital records of those collaborations opens

up the possibility of applying the proposed approach to larger document sources, such as corporate email.[2] Sandia is its human capital, and improving relationships in this domain will give us differentiated capabilities. Two, text-based instant messaging is used in other important customer domains, such as military and intelligence work (Ganter, 2007; Aoki, 2007). Strengthening our publicly releasable research credentials in these areas may provide additional means to attract funding from these customers. Finally, as mentioned above, a significant portion of the requested funding will be invested in the University of Texas strategic partnership, strengthening and deepening it.

Proposed Budget

The following budget has been designed to investigate the feasibility of the proposed technical idea at the estimated minimum cost possible.

The total budget requested is $35K, broken down as follows:

Funding for PI Scholand	$17K
Funding for University of Texas	$15K
Travel and Software	$3K

References

Aoki, P.M., "Back Stage on the Front Lines: Perspectives and Performance in the Combat Information Center," CHI 2007, April 28-May 3, 2007, San Jose, California, pp. 717-727.

Borgatti, S.P., Cross, R., "A Relational View of Information Seeking and Learning in Social Networks," *Management Science* Vol. 49, No. 4, (2003) pp. 432-445.

Cross, R., Parker, A.: *The Hidden Power of Social Networks*. Harvard Business School Press, Boston, MA 2004.

DiMicco, J.M., Yankelovich, N., "Constellation: Using Visualization to Find the Path to Experts," Group 07, ACM International Conference on Supporting Group Work, Sanibel Island, FL (2007).

Ehrlich, K., Lin, C.Y., Griffths-Fisher, V., "Searching for Experts in the Enterprise: Combining Text and Social Network Analysis," Group 07, ACM International Conference on Supporting Group Work, Sanibel Island, FL (2007) pp. 117-126.

[2] Based on conversations with Bruce Hendrickson, computations over this network would likely be appropriate for study under the FY07 Grand Challenge LDRD on Network Discovery, Prediction, and Disruption.

Fisher, D. and Dourish, P. (2004). "Social and Temporal Structures in Everyday Collaboration." Proceedings of the ACM Conference on Human Factors in Computing Systems CHI 2004 (Vienna, Austria), pp. 551-558.

Ganter, J. H. (2007). "Naturally Aligned: A Psycholinguistic Study of Shared Situation Awareness Among Military Sensor Operations Teams." Proceedings of the Eighth International NDM Conference (Eds. K. Mosier & U. Fischer), Pacific Grove, CA, June 2007.

Guimerà, R., Danon, L., Díaz-Guilera, A., Giralt, F., and Arenas, A. (2002). "Self-similar community structure in organizations." Physical Review E 68, 065103 (2003)

Hansen, M.: The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly* Vol. 44 (1999), pp. 82 -111.

Hu, Y., Fowler Wood, J., Smith, V., Westbrook, N., "Friendships through IM: Examining the Relationship between Instant Messaging and Intimacy." JCMC 10(1), Article 6, November 2004.

Lin, C.Y., Griffths-Fisher, V., Ehrlick, K., Desforges, C.: SmallBlue: People Mining for Expertise Search and Social Network Analysis, *IEEE Multimedia Magazine*, Oct – Dec 2007.

Tyler, J.R., Wilkinson, D.M., Huberman, B.A., "Email as Spectroscopy: Automated Discovery of Community Structure within Organizations, Communities, and Technologies," in Huysman, M., Wenger, E., and Wulf, V., Communities and Technologies, Kluwer BV 2003.

Biographical Sketch

Andrew J. Scholand received a B.S.M.E. with high distinction from Worcester Polytechnic Institute in 1989. After a year volunteering in Southern Africa installing solar panels, he obtained a Masters Degree in Electrical Engineering from Kings College London, graduating third in class. He returned to mechanical engineering for his Ph.D. at the Georgia Institute of Technology, focusing on analysis theory and methodology in computer aided engineering (CAE). At Georgia Tech, Andy held dual fellowships from the Department of Energy (Integrated Manufacturing Fellowship) and the state of Georgia (Georgia Tech President's Fellowship). Andy graduated from Georgia Tech in 2001, and joined Sandia National Laboratories 6 months later. He is currently researching the interaction of groups around complex technical software in 5535.