

Visualization of Interaction Patterns in Collaborative Knowledge Networks for Medical Applications

*Peter A. Gloor**

Scott Dynes

Yan.Zhao

Rob Laubacher

MIT CCS
Cambridge, MA USA
(pgloor, rjl}@mit.edu

Dartmouth Tuck CDS*
Hanover, NH USA
sdynes@acm.org

Dartmouth College
Hanover, NH USA
Yan.Zhao@dartmouth.edu

Abstract

By combining virtual communities with Internet portal and content management technologies, Collaborative Knowledge Networks (CKNs) (Gloor, 2002) share, access and extend the tacit and explicit knowledge within and across organizations. CKNs are a special kind of web-enabled communities of practice, where like-minded people collaborate and work together towards a common goal, sharing the same vision and values. CKNs are highly relevant also for biomedical Web-based communities to share knowledge and collaborate. In this paper we describe a system for the semi-automatic localization of CKNs in organizations. We identify structural properties and parameters of successful CKNs, based on automated analysis of e-mail archives. We then outline applications of CKNs in the medical field, looking at research collaboration such as new drug development, educational communities such as patient communities, and diagnosis and treatment communities such as clinical trial communities.

1 Introduction – What are CKNs

CKNs are nothing new. In fact, one of the most successful CKNs was started around 2000 years ago in the town of Jerusalem by a carpenter. But the fundamental principles are still the same: a truly innovative idea, which goes against conventional wisdom, is sold by a charismatic leader and a small group of dedicated disciples to an initially skeptical audience. The idea then catches on, and changes the way that the environment behaves.

For an eminent example, look at the way how the Web itself arose as a CKN, driven by visionary leaders: The community of early Web developers exhibited all the characteristics of a successful CKN at work, forming an intrinsically committed, dedicated community. Members joined out of their free will and collaborated not for immediate monetary gain, but because they shared the same values and beliefs. In the meantime the Web has become one of the main drivers of change for our economy, creating billions of dollars of wealth during this process. Even today thousands of dedicated volunteers work together in numerous CKNs to further drive the development of the Web.

Collaborative Knowledge Networks are a concept for entities that have always existed in and across organizations: groups of self-motivated individuals driven by the idea of something new and exciting - a way to greatly improve an existing business practice, or a new product or service for which they see a real need.

CKNs have obtained a great boost from the Internet by globally extending communication and collaboration facilities with instantaneous reach to anybody on the Internet. By combining virtual communities with Internet portal and content management technologies, CKNs can share, access and extend the tacit and explicit knowledge within and across organizations. In this way CKNs are

like communities of practice (Wenger, McDermott & Snyder, 2002), where people collaborate and work together towards a common goal. CKNs are distinct from communities of practice and other forms of online collaboration and virtual teams in both the informal nature of their work, and in the motivation for individuals to join and participate. In CKNs the motivation is internal – they do it because they are driven to do it, as opposed to participation being required as part of work or learning assignments as is often the case in virtual teams and communities of practice.

Due to this internal motivation CKNs cannot be mandated into action; in fact, ordering a CKN into existence is against the very foundation of how CKNs operate. People join and work in CKNs not because they have been told to do so by their superiors, but because they are personally motivated by and convinced of the vision and goals of the CKN community. CKNs are often spawned within an organization, but they quickly break organizational boundaries to include members from outside, bringing in new insights and knowledge otherwise not available within the organization. As noted above, CKNs involve individuals not necessarily related in terms of the corporate hierarchy.

CKNs are an important, unrecognized source of innovation within organizations; identifying and supporting CKNs should be a primary goal of every organization.

2 How to find CKNs

Just as Google is very effective at finding pertinent documents based on viewing patterns, we believe analysis of e-mail and other interaction logs of organizations will enable one to determine communities and core contributors. From the patterns and content of these interactions we will be able to create an index of the CKNs that exist within the organizations, as well as identify agents (individuals or groups) who are sources of expertise and also users of different classes of knowledge.

We are proposing a new methodology to trace the emergence of CKNs and their development over time, and to compute metrics of a CKN's efficiency and its implications for an organization's performance by mining computer logs such as email archives. Our proposed system computes and visualizes the structure of existing CKNs in organizations. We validate this methodology using information on the CKN collaboration structure obtained by interviews with CKN members and other assessments. To address the issues of the potential organizational impact of CKNs we will then compare the different types of CKNs with other corporate performance metrics such as new product output, share price, or profitability. This will allow us to draw conclusive evidence of the impact of CKNs on the performance of the organization.

3 Architecture of the CKN Visualization System

We have implemented a flexible three-level architecture (Figure 1). In the first step, the e-mail messages are parsed and stored in decomposed format in a SQL database. In the second step the database can be queried to select messages sent and/or received by a certain group in a certain time period. In the third step the selected communication flows can be visualized using SNA visualization tools such as Pajek (Batagelj & Mrvar, 1998) and ucinet (Borgatti, Everett & Freeman, 1992).

This architecture provides an optimal testbed of high scalability and flexibility: the number of messages to be analyzed is only limited by the size of the database, and temporal queries can be run in an ad hoc way. We will also be able to experiment with different visualizations of the

retrieved structure, identifying graphical representations that adequately reflect the temporal nature of the social networks.

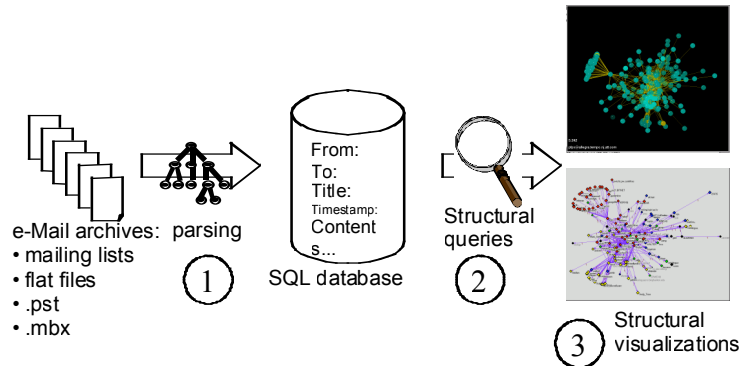


Figure 1: CKN e-mail analysis system architecture

4 First Results of CKN Visualization: KIF Community

As a proof of concept we have analyzed the e-Mail temporal communication flow of 555 e-mail messages contained in the <http://www.ksl.Stanford.EDU/email-archives/> of three mailing lists on shared reusable knowledge bases, the KIF knowledge interchange format, and on reusable ontologies. The messages come from about 220 active members. The “knowledge representation” CKN was active in the period between 1990 and 1994. For this initial analysis we examined the temporal flow of emails as a function of time. Figure 2 shows the email interactions for each year of the five years covered by the archive (1990 until 1994).

The goal of this initial analysis was to determine the feasibility of identifying the different roles of the community members of the CKN over time. The initial results are encouraging: in some years there are clearly individuals who are the main senders of emails, and others who are major recipients. It is interesting to note that these individuals change with time, and that the activity level of members of the CKN differs from year to year. This is particularly evident in 1990 and 1991, where there is very little overlap in the group of active members.

There is a rich literature about small group interactions, and we were curious as to whether the results of our initial analysis would conform to patterns identified in this literature. Comparing the activity over the lifetime of the “knowledge representation” CKN with Tuckman’s five classic stages of group development (Tuckman, 1965) (Forming, Storming, Norming, Performing, Adjourning), we can comfortably fit the activity patterns into this framework: 1990 is the forming year, where one active individual is initiating the CKN, while a second stream of scattered information is bypassing this most active member.

Storming involves a core team actively involved in determining the direction of the group; 1991 shows the storming phase with a core team interacting closely. 1992 and 1993 show the greatest activity of the CKN, bystanders of previous years now also become active contributors. This is consistent with Norming (selling the goals of the group to a wider audience and Performing (working towards the goals of the group). The last year, 1995, shows a decrease in the activity of the CKN, the activity level of the most active members degrades dramatically. This corresponds to the Adjourning stage.

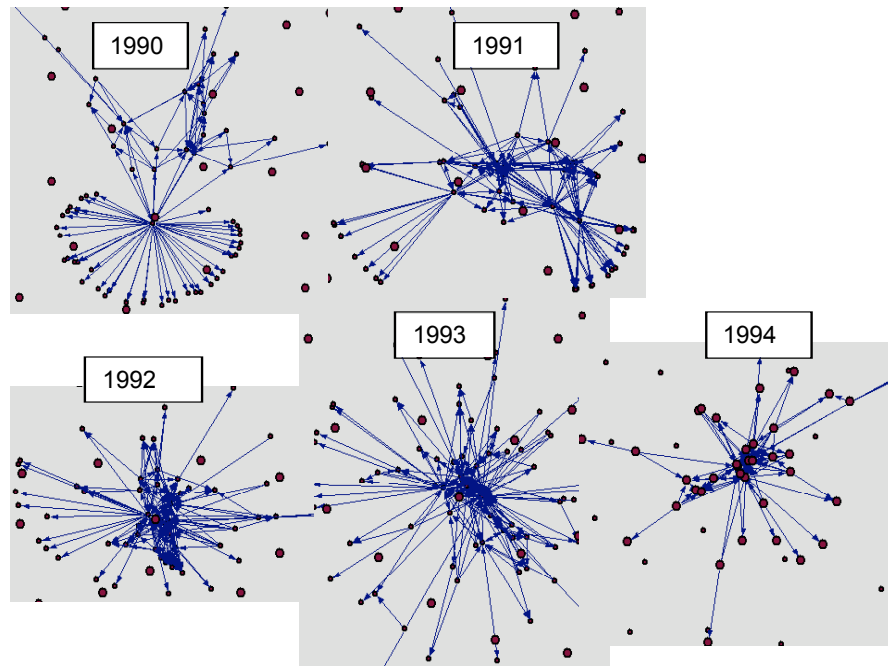


Figure 2: Email interactions as a function of time displayed in pajek. Each member is depicted as a dot. An arrow depicts email interaction between a pair of members.

5 Medical Applications of CKN

CKN concepts are of high practical relevance in the medical field. For example, they have been applied to analyze and improve the new drug development process for pharmaceutical companies (Gloor, 2002). Pharmaceutical companies are refocusing on core competencies and exploring new, more collaborative organizational models, as soaring R&D complexities and stricter regulatory standards have tripled R&D costs within the last 20 years. Companies are increasing and strengthening their collaborative relationships with external organizations for new product opportunities. Also, pharmaceutical decision-making now involves a network of influencers such as governments, insurers, HMOs, etc, although the main area of focus remains the physicians.

With growing levels of wealth and education, patients are becoming more health conscious and increase their influence in the drug buying process by turning to online shopping. For example, <http://www.healthtalkinteractive.com/> provides patients with interactive advice by experts and the opportunity to learn from the experience of their own patient community.

Applying CKN concepts internally leads to many advantages for pharmaceutical companies. First, they learn about innovations, which are happening in the company although senior management is not aware of it. This leads to better and more focused allocation of resources and to the identification of new business opportunities. Second, the company becomes more efficient in working together – thus dramatically cutting time to market for new inventions. Third, localization of CKNs helps to find knowledge faster. Visualization of CKNs allows to streamline communication processes, and to locate inefficiencies. Fourth, subject matter experts who might

not be high up in the corporate hierarchy can more easily be identified, allowing to better reward key contributors.

6 Conclusions

Social network analysis (SNA) researchers (Wassermann & Faust, 1994) have looked at knowledge networks for some time, but mostly by interviewing people and producing a snapshot view. (Girvan & Newman, 2001) computed the community structure of various popular SNA test data sets using a hierarchical clustering algorithm. (Guimera et al., 2002) computed the overall community structure of a university network by automatically analyzing the e-mail log. (Ebel, Mielsch & Bornholdt, 2002) analyzed the e-mail logs of Kiel University and found a small world networking structure with scale-free behavior (Barabasi, 2002). In our work we automatically analyze the e-mail logs of large organizations in order to extract and cluster the CKN sub-communities. We investigate large-scale networks over time, thus providing a unique temporal view on social networks of thousand of people.

Our research will further the understanding of how virtual teams are formed, how they function, and how they die. The knowledge and tools developed will allow medical researchers, managers, and politicians to better understand how to find CKNs, how they function, what drives members to join, what the crucial roles in CKNs are, and how their success can be measured.

7 Acknowledgements

We are grateful to Tom Malone, Hans Brechbuehl, John Quimby, Tom Allen, M. Eric Johnson, JoAnne Yates, Fillia Makedon, and Jue Wang for their help and encouragement. The Dartmouth DEVLAB provided the IT infrastructure.

8 References

- Barabasi, (2002). A. Linked: the new science of networks. Perseus, Cambridge, MA.
- Batagelj, V. Mrvar, A. (1998). Pajek - Program for Large Network Analysis. Connections 21, 2, 47-57.
- Borgatti, S., Everett, M. Freeman, L.C. (1992). UCINET IV, Version 1.0, Columbia: Analytic Technologies.
- Ebel, H. Mielsch, L. Bornholdt, (2002). S. Scale-free topology of e-mail networks. arXiv:cond-mat/0201476v2 12 Feb 2002.
- Girvan, M. Newman, M.E.J. (2001). Community structure in social and biological networks. arXiv:cond-mat/0112110v1, 7 Dec 2001.
- Gloor, P. (2002). Collaborative Knowledge Networks. eJETA the electronic Journal for e-Commerce Tools and Applications, vol 1, no. 2, November. www.ejeta.org.
- Guimera, R., Danon, L., Diaz-Guilera, A. Giralt, F., Arenas, A. (2002). Self-similar community structure in organizations. ArXiv:cond-mat/0211498 v1, 22 Nov 2002.
- Tuckman, B.W. (1965). Developmental Sequence in Small Groups, Psychological Bulletin, 63, 384-99.
- Wasserman, S., Faust, K. (1994). *Social Network Analysis : Methods and Applications*. Cambridge University Press.
- Wenger, E. McDermott, R. Snyder, W. (2002). *Cultivating Communities of Practice*. Harvard Business School Press.