

## Association for Information Systems AIS Electronic Library (AISeL)

---

ICIS 2003 Proceedings

International Conference on Information Systems  
(ICIS)

---

December 2003

# Informal Knowledge Networks: Toward a Community-Engineering Framework

Gregor Schrott  
*Frankfurt University*

Daniel Beimborn  
*Frankfurt University*

Follow this and additional works at: <http://aisel.aisnet.org/icis2003>

---

### Recommended Citation

Schrott, Gregor and Beimborn, Daniel, "Informal Knowledge Networks: Toward a Community-Engineering Framework" (2003). *ICIS 2003 Proceedings*. 77.  
<http://aisel.aisnet.org/icis2003/77>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2003 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# INFORMAL KNOWLEDGE NETWORKS: TOWARD A COMMUNITY-ENGINEERING FRAMEWORK

**Gregor Schrott and Daniel Beimborn**

Institute of Information Systems

Frankfurt University

Frankfurt, Germany

[gschrott@wiwi.uni-frankfurt.de](mailto:gschrott@wiwi.uni-frankfurt.de)

[beimborn@wiwi.uni-frankfurt.de](mailto:beimborn@wiwi.uni-frankfurt.de)

## Abstract

*The problems knowledge workers face today are dynamic, unstructured, highly complex, and often cannot be fully explicated. Such moving targets require different problem solving capabilities. Because abstract information is less valuable in this type of environment, knowledge workers have to utilize channels other than handbooks. Hence, corporate knowledge networks again are at the top of the research agenda. For a knowledge worker, access to knowledgeable colleagues, rather than access to large databases, becomes the important factor.*

*In such networks, the question of which managerial actions are appropriate for successful community development (i.e., supporting the actors) arises. Unfortunately, today's community engineering practices are often characterized by a gross simplification and strong technological focus rather than modeling the impact of managerial actions before taking them.*

*As part of a larger research project, this paper addresses topological structures as an action variable of community engineering. A computer-based simulation model is introduced and applied to real-life data from over 800 students and staff of the Economics and Business Administration Department at Frankfurt University, Germany.*

## Introduction

The increasing importance of corporate knowledge networks is leading to a rebirth of the main challenges of knowledge management: adequate support of the actors, teams, and communities in virtual knowledge networks within corporations (Davenport and Prusak 1998; Wenger 2000). In this work, knowledge networks are understood as “interconnected nodes (individuals) linked by arcs or edges (communication flows), representing information communication patterns that crystallize over time” (Ahuja and Carley 1999). These informal networks are very important to employees, especially in fast-changing organizational environments (Nardi et al. 2000) and have been recognized by many corporations (Botkin 1999; Enkel et al. 2000, Lesser et al. 2001; Lesser and Prusak 1999). Much has been written about knowledge management and virtual communities in general, but only part of the literature has inspired applicable results for the management of and practice in virtual knowledge networks (Ahuja and Carley 1999). Practices often are based on inaccurate perceptions of the organization's informal structure, characterized by random interventions rather than systematic network development (Cross et al. 2001; Krackhardt and Hanson 1993) and a strong focus on technical aspects. We believe that the communication structures underlying knowledge networks are a key determinant influencing the efficiency of information supply. Therefore, knowledge officers should consider communication topology as endogenous. Methods of social network analysis allow for the precise capturing of such network structures (Wassermann and Faust 1994), and thereby act as a first step toward the development of appropriate solution strategies. But being able to generate highly accurate “snapshots” of structures is not sufficient for successful community support. Managers

responsible for network support should be able to estimate the effects caused by their actions before applying them to sensitive virtual knowledge networks (Cross and Prusak 2002). One goal of our work is to develop and evaluate solution strategies for corporate knowledge officers facing decisions on community-engineering activities in order to improve the information supply within virtual knowledge networks. As a first step, we introduce a basic model for analyzing the effect of a simple topological change action on the performance of a network.

The research question is: which information topology is best suited for successful knowledge networks? This paper is organized as follows: based on a short literature review, we develop a new framework for testing the effect of community-engineering activities. We then apply our model to empirical data collected during the summer of 2002 at the Economics and Business Administration Department at Frankfurt University, Germany. The findings of the computer simulations are presented and discussed. The findings are then summarized and our next research steps described.

## Related Work

The field of Computational Organization Theory (COT) (Carley and Prietula 1994) studies organizational phenomena by using autonomous intelligent agents. As articulated by Monge and Contractor (2003, “the agent-based approach to the study of complex systems is especially suited to understand knowledge networks,” because COT allows for the study of highly complex phenomena. In this work, all actions that aim at actively influencing the network topology will be subsumed under the term *community engineering*. Depending on the domain and level of interference, examples include the design of office spaces (coffee-corners) or job rotation programs increasing the number of communication links within the network (people get to know new people) and closed virtual workspaces and weekly team meetings intensifying existing links within the network (people get to know each other better).

## A Simulation Model for Virtual Knowledge Communities

### Network Model

The modeled network consists of actors (knowledge workers)  $i \in NW$  with  $i = 1 \dots I$  (NW: set of actors) and relations between these actors. The relations are modeled by a relationship vector  $\bar{e}_i$  for every actor  $i$ :

$$\bar{e}_i = (e_{i1} \dots e_{ij} \dots e_{iI}) \text{ with } e_{ij} = \begin{cases} g_{ij} & \text{if } j \in NH_i \\ 0 & \text{else} \end{cases} \quad (1)$$

If actor  $i$  has a link to actor  $j$ ,  $j$  is called “neighbor of  $i$ .”  $NH_i$  represents the set of actor  $i$ 's neighbors ( $NH_i \subseteq NW \quad \forall i \in NW$ ). In case  $i$  has a link to actor  $j$ ,  $e_{ij}$  is set to  $g_{ij}$ , which represents the edge weight, hence the intensity of their relation. A positive value indicates that actor  $i$  knows actor  $j$  and is able to ask him for advice concerning his problems. Every actor has certain knowledge, i.e. he holds certain information objects. His personal information pool is defined as the knowledge base  $K_i$  and is represented by the knowledge vector  $\bar{o}_i$ :

$$\bar{o}_i = (o_{i1} \dots o_{ik} \dots o_{iK}) \text{ with } o_{ik} = \begin{cases} 1 & \text{if } k \in K_i \\ 0 & \text{else} \end{cases} \quad (2)$$

The knowledge vector consists of all  $K$  different information objects, which exist in the network and are distributed over all actors. If actor  $i$  holds object  $k$  (knows  $k$ ), then the binary variable  $o_{ik}$  is 1.  $K$  is the total number of different existing information objects in the network.

**Problems/information needs:** In the network, actors will face several problems over time (information needs arise) which have to be solved within a given timeframe  $TP$ . After this time (deadline), the information objects are of no value to the actor (Barreau and Nardi 1995; Kidd 1994). For our model, this implies that the status of the problems becomes persistent. For simplification, we assume the same timeframe  $TP$  for all actors and problems and the problem sets of the different actors are disjoint. Hence, actor  $i$ 's actual set of problems  $P_{it}$  is modeled as problem vector  $\bar{p}_{it}$ :

$$\bar{p}_{it} = (p_{i1t} \dots p_{ikt} \dots p_{iKt}) \text{ with } p_{ikt} = \begin{cases} 1 & \text{if } k \in P_{it} \\ 0 & \text{else} \end{cases} \quad (3)$$

$p_{ikt}$  again is a binary variable and is equal to 1 if actor  $i$  has problem  $k$  in period  $t$ . Every emerging problem has a corresponding information object actors do not know about, i.e., they don't know where the respective information object exists; they just know that there is a potential solution for their problem. Finding this object in the network solves the problem. If  $p_{ikt}$  is equal to 1, the problem could have emerged in period  $t$  or earlier if it has not been solved yet. The problems remain active until being solved:

$$p_{ikt} \leq p_{ikt+1}.$$

**Requests:** When searching for the information objects, actors send requests to their neighbors in order to find the answer to their problems. These neighbors in turn forward the request to their neighbors, and so on. The actor-based communication acts per period are represented as follows:

Sending requests:

$$R_{it}^+ = \begin{pmatrix} r_{i11t}^+ & \dots & r_{ij1t}^+ & \dots & r_{i11t}^+ \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{i1kt}^+ & \dots & r_{ijkt}^+ & \dots & r_{i1kt}^+ \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ r_{i1Kt}^+ & \dots & r_{ijKt}^+ & \dots & r_{i1Kt}^+ \end{pmatrix} \forall i; t \quad (4)$$

with  $r_{ijkt}^+ = \begin{cases} 1 & \text{if } i \text{ sends request to } j \text{ about } k \text{ in } t \\ 0 & \text{else} \end{cases}$

Receiving requests is modeled accordingly, with opposite sign  $r_{ijkt}^-$ .

**Restrictions:** The actors in the network have a certain number of maximal requests they are willing to serve. This restriction represents the number of e-mails someone will process during one period (e.g., a day) of the simulation. Due to limited availability as well as willingness to invest in other peoples' problems, this restriction exists. Therefore, we define MC as the general maximal capacity of accepted requests (messages):

$$\sum_{j \in NH_i} \sum_{k=1}^K r_{ijkt}^- \leq MC \quad \forall t \quad (5)$$

Despite actors' different willingness and ability to help within virtual knowledge networks (Nonnecke and Preece 2003), for simplification purposes we assume that MC is equal for all actors. To determine the incoming messages that are forwarded, we use a recognition probability  $pr[r_{ijkt}^-]$  that a message will be recognized by actor  $i$ . The probability depends on four things. First, it depends on the total number of messages that have been sent to the actor within period  $t$ . The higher the total number of messages received by actor  $i$  (c.p.), the less likely a single message will be recognized. Second, it depends on the weight of the respective edge  $e_{ij}$  from sender to receiver. The more intense actor  $j$ 's relationship with actor  $i$  the more likely actor  $i$  will recognize his request. Third, it depends on the number of requests asking for the same information object. The more people ask for the same information object, the more likely the message will be recognized. Fourth, it depends on the number of messages that are recognized ( $MC$ ). Hence, good relationships as well as multiple requests for the same information object increase the expected likelihood of messages being recognized.

$$Exp[r_{ijkt}^-] = pr[r_{ijkt}^-] \cdot 1 = \frac{MC \cdot \sum_{j \in NH} r_{jikt}^+ \cdot e_{ji}}{\sum_{k=1}^K \sum_{j \in NH} r_{jikt}^+ \cdot e_{ji}} \quad \forall i, k, t \quad (6)$$

**Problem Solving Act:** The problem  $k$  of an actor  $h$  will be solved if one of his messages which could be forwarded by any other actor  $i \in NW$  will be received by the actor  $j$  who holds the corresponding information object.

$$\left( \sum_{\substack{j \in NH_i \\ j \neq i}} r_{ijkt} \geq 1 \wedge o_{ik} = 1 \text{ for at least one actor } j \right) \text{ then } \Rightarrow p_{hkt+1} = 0$$

$$\text{while } \sum_{j \in NH_i} \sum_{k \in P} r_{ijkt}^- \leq MC$$

**Decision Process:** Each actor has to decide twice each period. He needs to decide, first, about forwarding incoming messages and, second, about sending his own requests for solving his own problems. When receiving requests, actor  $i$  either holds the information required and is able to solve the problem (see above), or he does not have the information required and has to decide about forwarding requests. The decision variable for actor  $i$  is  $r_{ijkt}^+$  representing the decision about relaying received messages. Which neighbors will get the forwarded message? Actor  $i$  selects  $Z$  neighbours (or if he has less, then all neighbors) by an equally distributed random process incorporating the edge weight (intensity of the relationship). The more intense his relation with an actor  $j$ , the more likely he will ask that particular actor. The probability that actor  $j$  receives a request from actor  $i$  concerning object  $k$  in period  $t$ , represented as expectation of  $r_{ijkt}^+$ , is calculated as follows:

$$Exp[r_{ijkt}^+] = \frac{e_{ij} \cdot \min\left(Z, \sum_{j \in RNH_{it}} sign(e_{ij})\right)}{\sum_{j \in RNH_{it}} e_{ij}} \quad \forall i, j, k, t \quad (7)$$

To ensure that exactly  $Z$  neighbors will be asked, the following restriction has to be met:

$$\sum_j r_{ijkt}^+ = \min\left(Z, \sum_{j \in RNH_{it}} sign(e_{ij})\right) \quad \forall i, k, t \quad (8)$$

$RNH_{it}$  indicates the subset of neighbors of actor  $i$  that are relevant in terms of being able to receive messages in period  $t$ .  $RNH_{it}$  does not include those actors  $j$  having asked actor  $i$  for the information object  $k$  until the current period and those actors  $j$ , which have been asked by actor  $i$  for the information object  $k$  in the past:

$$RNH_{it} \subset NH_i \subseteq NW \quad \forall i \in NW \quad (9)$$

**Sending Requests:** Choosing the neighbor for sending requests to solve his own problems follows the same rationale as forwarding requests. For every active problem (max.),  $Z$  neighbors will be asked.

**Measurement of Network Efficiency:** Finding the appropriate efficiency concepts of network design can follow different paths, depending on the context of the problem-solving network. In this model, we take a straightforward approach to measure the effect of different network designs based on Carley and Lin (1997). The individual knowledge coefficient measures the problem solving efficiency by building a ratio of solved problems and the total number of problems that appeared for each actor  $i$ :

$$KQ_i = \sum_{t=1}^T \sum_{k \in K_i} \frac{x_{ikt}^+}{x_{ikt}^-} \quad \text{with } x_{ikt} = p_{ikt-1} - p_{ikt} \quad (10)$$

where  $x_{kt}^+$  represents all strictly positive values of  $x_{kt}$  (all problems being solved) and  $x_{kt}^-$  represents all negative values (appeared problems). The knowledge coefficient ranges from 1 (all problems solved) to 0 (no problem solved). Besides the individual measure of network efficiency, the overall knowledge coefficient can be calculated as follows:

$$KQ = \frac{\sum_{i \in NW} KQ_i}{I} \quad (11)$$

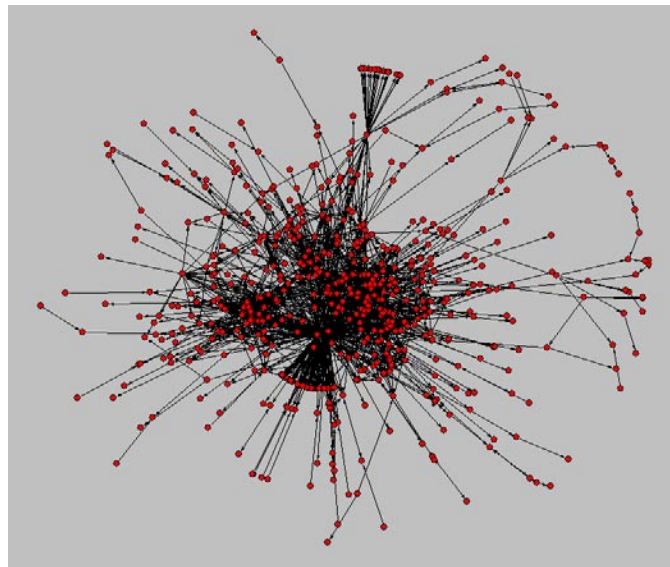
While the knowledge coefficients are a good measure of overall performance of the network, they do not measure how efficiently the problems have been solved. Therefore, we define SH as search hops per solved problem:

$$SH = \frac{\sum_{t=1}^T \sum_{i=1}^I \sum_{\substack{j=1 \\ j \neq i}}^I \sum_{k=1}^K r_{ijkt}^+}{\sum_{t=1}^T \sum_{k \in K_i} x_{ikt}^+} \quad (12)$$

where  $r_{ijkt}^+$  represents all strictly positive values of  $r_{ijkt}$ .

### Simulation Design

All simulation results were generated using JAVA 1.4 applications. For data analysis PAJEK (Batagelj and Mrvar 1998), SPSS, and MS Excel were used. To model an exemplary network, we used e-mail exchanges from over 800 members of the Faculty of Economics and Business Administration at Frankfurt University, Germany. During August 2002, over 3,000 e-mail messages were collected. The network (cf. Figure 1) consists of 808 actors, connected by 2,573 relations. The relations are directed (arcs) and indicate the total number of messages send from actor  $i$  to actor  $j$  within the four-week period. Hence, the network is very sparse. The density (number of existing links/number of possible links) is very low at 0.0039.



**Figure 1. Initial E-Mail Network**

Based on this specific network, we ran simulations of problem solving behavior (50 runs per parameter constellation). Limiting the number of incoming messages (MC) to 20 and assuming a maximum problem age (TP) of 30 periods,  $Z$  was varied from 1 to 10. Each network was simulated for 100 periods. The following table summarizes the initial parameter values.

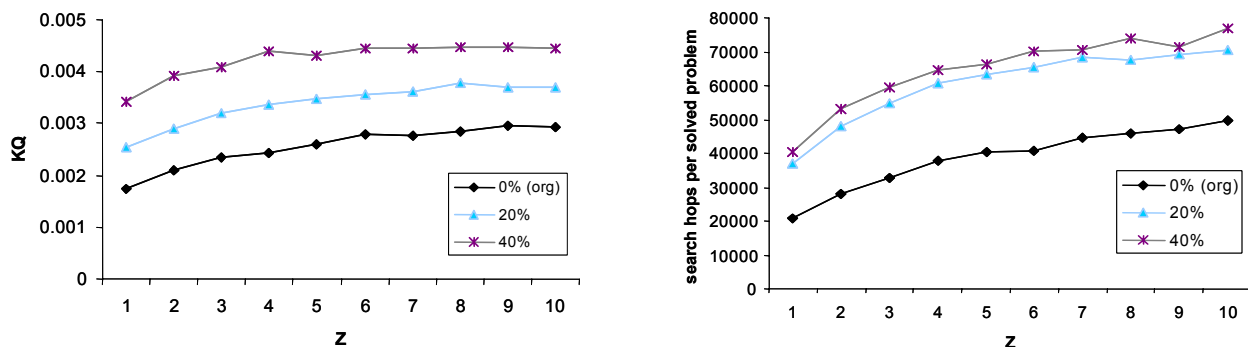
**Table 1. Parameterization**

Capacity of incoming messages (MC)	20
Avg. number of problems appearing per actor and period (PC)	0.3
Number of maximum recipients of the same message (Z)	Var.
Simulation horizon (T)	100
Max. age of a problem (TP)	30

To show the applicability of the proposed model, we use one action of community engineering: increasing the *network density* (number of existing links/number of possible links), which means that people get to know more other actors of the network. In real-world terms, this could be stimulated by introducing new research assistants to the whole faculty via mailing lists. We increased the density of the original network (0.0039) about 20 percent (0.0043) and 40 percent (0.0055).

## Simulation Results

Increasing the *network density* leads to the effect displayed in Figure 2.



**Figure 2. Network Efficiency (KQ)/SH per Solved Problem/Increasing Density**

As can be seen from Figure 2, increasing network density (overall Z) leads to a higher number of problems being solved in the network (KQ). Increasing the number of links by about 40 percent leads to an average increase of solved problems of 68 percent. But increasing the number of messages probably comes at a high price: When considering the number of messages in relation to the number of solved problems (the right-hand side of Figure 2), it can be seen that with increasing density, more messages are being sent through the network. Hence, the advantages of more problems being solved have to be evaluated by knowledge officers against the costs of message handling of the network members.

## Summary and Further Research

The design of corporate knowledge networks is not purely a technical question. Unfortunately, little is known about the dynamics of processes within those networks and especially how to measure the effect of community-engineering actions undertaken by corporate knowledge officers. Our contribution to this problem is the attempt of the development of a universal simulation toolbox for knowledge networks which enables managers in charge of the network to anticipate the effects of their actions especially when facing cost-intense and long-term decisions. As a first step, in this paper a basic model was introduced and applied to exemplary real-world data. This paper does not aim at designing and evaluating different alternative actions of community engineering but introduces a model to measure the effect of various actions of community engineering. Of course, this model can only be seen as a first step in a research framework analyzing the measurements of such actions. Most of our assumptions are quite restrictive and will be relaxed subsequently. In addition, the current model will be extended to incorporate these dynamics.

## Acknowledgements

This research is supported by the German National Science Foundation (Ph.D. Program 492, “Enabling Technologies for Electronic Commerce”). We gratefully acknowledge the financial support. Also we would like to thank Tim Weitzel for his insightful comments on earlier versions of this paper and Wolfgang König for providing such a great academic home and encouraging us to work on this topic.

## References

- Ahuja, M., and Carley, K. M. “Network Structure in Virtual Organizations,” *Organization Science* (27:6), 1999, pp. 741-757.
- Batagelj, V., and Mrvar, A. “PAJEK-Program for Large Network Analysis,” *Connections* (21), 1998, pp. 47-57.
- Barreau, D., and Nardi, B. “Finding and Reminding: Organization of Information from the Desktop,” *SIGGHI Bulletin* (27), 1995, pp. 39-45.
- Botkin, J. *Smart Business: How Knowledge Communities Can Revolutionize Your Company*, Free Press, Boston, 1999.
- Carley, K. M., and Lin, Z. “A Theoretical Study of Organizational Performance Under Information Distortion,” *Management Science* (43:7), 1997, pp. 976-997.
- Carley, K. M., and Prietula, M. J. (eds.). *Computational Organization Theory*, Lawrence Erlbaum, Hillsdale, NJ, 1994.
- Cross, R., Parker, A., Prusak, L., and Borgatti, S. P. “Knowing What We Know: Supporting Knowledge Creation and Sharing in Social Networks,” *Organizational Dynamics* (30:2), 2001, pp. 100-120.
- Cross, R., and Prusak, L. “The People Who Make Organizations Go—or Stop,” *Harvard Business Review*, June, 2002, pp. 5-12.
- Davenport, T., and Prusak, L. *Working Knowledge*, Harvard Business School Press, Boston, 1998.
- Enkel, E., Heinold, P., Hofer-Alfeis, J., and Wicki, Y. “The Power of Communities: How to Build Knowledge Management on a Corporate Level Using the Bottom Up Approach,” in *Knowledge Management Case Book*, T. Davenport and G. Probst (eds.), John Wiley & Sons, New York, 2000, pp. 84-103.
- Kidd, A. “The Marks Are on the Knowledge Worker,” in *Proceedings of the CHI’94 Human Factors in Computing Systems*, Boston, 1994, pp. 186-191.
- Krackhardt, D., and Hanson, J. R. “Informal Networks: The Company Behind the Chart,” *Harvard Business Review*, July/August 1993, pp. 104-111.
- Kristoffersen, S., and Ljungberg, F. *Your Mobile Computer Is a Stationary Computer*, Working Paper No. VRR-98-34, Viktoria Institute, Goteborg, Sweden, 1998.
- Lesser, E. L., Fontaine M. A., and Slusher, J. A. (eds.). *Knowledge and Communities*, Butterworth Heinemann, Boston, 2001.
- Lesser, E., and Prusak, L. *Communities of Practice, Social Capital and Organizational Knowledge*, Working Paper, IBM Institute for Knowledge Management, Boston, 1999.
- Monge, P. R., and Contractor, N. *Theories of Communication Networks*, Oxford University Press, New York, 2003.
- Nardi, B., Whittaker, S., and Schwarz, H. “It’s Not What You Know, It’s Who You Know: Work in the Information Age,” *First Monday*(5:1), 2000 (available online at <http://www.firstmonday.com>; accessed May 1, 2003).
- Nonnecke, B., and Preece, J. “Silent Participants: Getting to Know Lurkers Better,” in *From Usenet to CoWebs—Interacting with Social Information Spaces*, C. Lueg and D. Fisher (eds.), Springer, Berlin, 2003, pp. 110-133.
- Wasserman, S., and Faust, K. *Social Network Analysis*, Cambridge University Press, Cambridge, UK, 1994.
- Wenger, E. “Communities of Practice: The Key to Knowledge Strategy,” in *Knowledge and Communities*, E. Lesser, M. Fontaine, and J. Slusher (eds.), Butterworth, Boston, 2000.