A DYNAMIC NETWORK ANALYSIS APPROACH FOR EVALUATING KNOWLEDGE DISSEMINATION IN A MULTI-DISCIPLINARY COLLABORATION NETWORK IN OBESITY RESEARCH

Hyojung Kang
David Munoz

The Pennsylvania State University
Department of Industrial and Manufacturing Engineering
362 Leonhard Building
University Park, PA 36849, USA

ABSTRACT

Effective knowledge dissemination is important to promote the adoption of new concepts and tools. This study aims to provide a framework that assesses strategies for successful knowledge dissemination in a research collaboration network. We propose a Markov-chain Monte Carlo (MCMC) approach along with Dynamic Network Analysis (DNA) to model a social network and understand how different knowledge dissemination strategies can be used in a research collaboration network. The proposed method was demonstrated through a case study that uses a multi-disciplinary collaboration network in obesity research at an academic medical center. To assess the impact of initial disseminators on knowledge dissemination, four different strategies were considered. The simulation results indicated that the best strategy to disseminate knowledge within this obesity research network may be to use central agents in clusters when considering the coverage and speed of knowledge dissemination.

1 INTRODUCTION

Knowledge diffusion or dissemination has been defined as the process and extent of information exchange within an organization (Van Der Bij, Song, and Weggeman 2003). This communication process can occur through formal or informal channels. Moreover, it can be transferred under a horizontal or vertical organizational structure. Furthermore, the definition of knowledge translation can be used at the conceptual level (learning or acquisition of new perspectives and attitudes) or instrumental level (modified new practices) (Hutchinson and Huberman 1994). Within social systems, knowledge diffusion can be seen as the spontaneous or conscious effort to spread new knowledge, ideas, policies, and practices.

Knowledge diffusion within and across organizations is important in various areas ranging from industry to academia because it increases awareness of new concepts, practices, and tools, which leads to developing innovative ideas (Green et al. 2009, Ernst and Kim 2002, Azoulay, Zivin, and Sampat 2011). However, transferring knowledge between social agents has been considered to be a challenging task (Ward, House, and Hamer 2009).

In medicine, many studies have been conducted to develop evidence-based clinical guidelines to help clinicians make appropriate healthcare decisions and as a result to improve quality of care. However, the evidence and guidelines have been implemented only slowly in practice (Titler 2008, National Health and Medical Research Council 2000). One of the barriers to the limited use of the guidelines in healthcare practice includes clinicians’ lack of awareness of practice guidelines and resistance to changes (Pierson 2009).
Effective knowledge dissemination about new guidelines among clinicians in similar fields may overcome the barrier and facilitate the transfer of new evidences into practice.

This study aims to provide a framework that assesses strategies for successful knowledge dissemination in a research collaboration network. We proposed a Markov-chain Monte Carlo (MCMC) approach along with Dynamic network analysis (DNA) to model the network and understand how different knowledge dissemination strategies can be used in a research collaboration network. The proposed method was demonstrated through a case study that uses a multi-disciplinary collaboration network in obesity research at an academic medical center.

2 BACKGROUND

2.1 Dynamic Network Analysis

One of the main advantages of DNA over traditional social network analysis (SNA) is that it incorporates the dynamic nature of social networks. In this sense, it supports the modeling of networks under a more comprehensive structure that allows the prediction of important features such as knowledge diffusion. According to Carley (2003), there are three key advantages of using DNA: 1) combining topics of management, operations research, and social networks (meta-matrix), 2) using probabilistic ties between agents, and 3) combining traditional SNA with cognitive science and multi-agent systems.

Although DNA is an emergent field, it has been used in a wide range of applications including the assessment of terrorist groups (Carley 2006), evaluating social influence (Robins, Pattison, and Elliott 2001; Hill and Carley 2008), detecting change in human behavior (McCulloh and Carley 2008) and social networks (McCulloh and Carley 2011), and diffusion of spam through an email network (Mezzour and Carley 2014).

2.2 Knowledge Diffusion in Social Networks

In the past few years, increasing attention has been paid to the concept of knowledge diffusion within a social network. The main aim of the knowledge created at one point in time is to reach a certain space where the knowledge is needed at the time it is needed. Studies have investigated key factors that facilitate knowledge transfer such as people, organizational structure and culture, leadership, and information system (Susanty, Handayani, and Henrawan, 2012). Of these factors, people including the source and the recipient of information have a significant impact on successful knowledge transfer in a network. Both the capability of the source to provide necessary knowledge and the ability of the recipient to absorb the transferred knowledge affect the effectiveness of knowledge transfer (Argote and Ingram 2000; Al-Salti and Hackney 2011). The degree of interaction between the parties and the degree of similar knowledge bases they share are also important for adequate knowledge transfer (Cummings and Teng, 2003).

The capability of a network to influence its actors is associated with the rate at which knowledge is diffused and absorbed in the network. Therefore, it becomes more important to understand and identify enhanced strategies to transmit information based on current relationships or patterns of social networks. Some of these strategies are framed under knowledge brokering theory that seeks to facilitate the spread of knowledge within an organization through different key agents or brokers within a network (Ward, House, and Hamer 2009). Oldham and McLean (1997) propose three frameworks for thinking about these concepts to cover the creation, diffusion, and use of knowledge, as well as the generation of “creators” to foster links between agents, and enhancing access to knowledge through training to specific agents that may lead to positive social outcomes. In this study, we propose a DNA model for assessing different knowledge diffusion strategies in a multi-disciplinary collaboration network.
3 METHODOLOGY

This study developed and evaluated a DNA model using an MCMC approach. This approach considered a static network in terms of its structure (same agents and edges), but dynamic in the status of the information flow and parameters that approximate the knowledge transference over time.

Matrix $A$ represents a network $G: (V,E)$, in which $V$ is a set of nodes and $E$ is a set of paired nodes or edges ($a_{ij}$):

$$
A = \begin{bmatrix}
V_1 & V_2 & V_3 & \ldots & V_N \\
V_1 & a_{12} & a_{13} & \ldots & a_{1N} \\
V_2 & a_{21} & - & a_{23} & \ldots & a_{2N} \\
V_3 & a_{31} & a_{32} & - & \ldots & a_{3N} \\
\vdots & \vdots & \vdots & \vdots & - & \vdots \\
V_N & a_{N1} & a_{N2} & a_{N3} & \ldots & -
\end{bmatrix}
$$

Here, $a_{ij}$ represents a certain degree of connection between nodes $i$ and $j$. In a knowledge dissemination context, this parameter ($a_{ij}$) could represent the intensity of collaboration between two individuals. If $a_{ij}=0$, there is no link between $i$ and $j$. In this case, information cannot be disseminated directly between nodes $i$ and $j$. However, if the nodes have other “neighbors” in common, indirect knowledge dissemination between the nodes is possible.

The DNA model is supported by dynamic rules to approximate agents’ behavior and knowledge transference parameters. The rules governing these parameters are arranged into four matrix structures that are updated at each step $t$. These four probabilistic matrices are: 1) probability of meeting during a certain period $t$ (matrix $M$); 2) level of knowledge transference at a certain period $t$ (matrix $P$); 3) cumulative conversion level up to period $t$ (matrix $C$); and 4) agents converted up to period $t$ (matrix $K$).

Probability of meeting during a certain period $t$

In a social context, it is necessary that two agents communicate to disseminate knowledge from one agent to another agent. This communication process could be conducted through different communication channels and mechanisms, from face-to-face meetings to social media interactions.

A matrix $M$ represents the probability that agents $i$ and $j$ communicate during a certain period $t$ that can be represented by days, weeks, months, etc. A matrix $M^*$ shows whether the two agents communicate ($m_{ij}^* = 1$) or not ($m_{ij}^* = 0$) during the same period. These two matrices can be expressed as a symmetric network as the probability of $i$ meeting $j$ is the same as the probability of $j$ meeting $i$.

The probability of $i$ and $j$ to meet during a certain period $t$ ($m_{ij}$) is based on the frequency of connection between nodes ($a_{ij}$). Communications between researchers who have only a small number of collaborative projects may be less frequent, compared to communications between those who have intensively collaborated. Considering different rates of change in the communication probability based on their previous collaboration experiences, it was assumed that $m_{ij}$ is determined by a log function of $a_{ij}$.

$$
m_{ij} = f(a_{ij}) = \varphi \ln(\beta a_{ij}) \quad (1)
$$

Here, $\varphi$ and $\beta$ can be derived by assuming two data points, $P_1(x_1, y_1)$ and $P_2(x_2, y_2)$ where $x_1$ and $x_2$ represent the strength of connection between two agents ($a_{ij}$) and $y_1$ and $y_2$ represent the corresponding probability of meeting during a certain period ($m_{ij}$). Hence, by setting the two points, the parameters $\varphi$ and $\beta$ can be calculated as $\varphi = (x_2 - x_1)/\ln(y_1/y_2)$ and $\beta = \exp[(x_2 \ln y_1 - x_1 \ln y_2)/(x_1 - x_2)]$. For this study, it was assumed that new connections among nodes are not possible, hence, the probabilities $m_{ij}$ remain the same for each period $t$. To determine $m_{ij}^*$ at $t$, $m_{ij}$ is compared against a random number ($r$), which is generated from a uniform distribution between 0 and 1. A new binary variable $m_{ij}^*$ is calculated at each
period $t$. If $m_{ij} > r$, then $m_{ij}^* = 1$, otherwise $m_{ij}^* = 0$. Similar approaches have been used to determine the probability of meeting or interacting (Canals, Boisot, and MacMillan, 2005).

The level of knowledge transfer at a certain period $t$

We refer to an agent as converted when this agent has received and adopted new disseminated knowledge. This agent converted at period $t$ is able to convince and disseminate the knowledge to other agents from period $t+1$. As indicated in the previous section (2.2), there are various factors affecting the extent of knowledge dissemination between agents. Among the factors, we focused on the capability of the sender and the extent of interaction between the sender and the recipient. First, the capability of the sender was represented by sender’s reputation within the network ($a_{ii}/\max_i a_{ii}$). Here, it was assumed that the number of publications in a certain field is associated with the investigator’s capability to understand emerging knowledge in the field and transfer the knowledge to a network. For instance, an experienced investigator would have a greater influence on disseminating new knowledge than a novice investigator. Secondly, the degree of interaction between the sender $i$ and the receiver $j$ was represented by the proportion of collaborative work between them to the receiver’s research accomplishment ($a_{ij}/a_{jj}$). It was assumed that as the extent of collaboration between investigators increases, their relationship becomes stronger, which leads to smoother knowledge transition. Considering these two factors, a matrix $P$ representing the level of knowledge transfer is constructed as follows:

$$p_{ij,t} = k_{i,t}(1 - k_{j,t})m_{ij}^* \frac{\delta a_{ij}}{\max_i a_{ii}} + \frac{1 - \delta}{a_{jj}}$$  \hspace{1cm} (2)

where $p_{ij,t}$ is the degree of knowledge transfer from agent $i$ to $j$ at time $t$, in which $p_{ij,t}=0$ means no knowledge transferred from $i$ to $j$ and $p_{ij,t}=1$ means the full amount of knowledge is transferred from $i$ to $j$. $k_{i,t}$ is an indicator variable that shows the status of agents’ knowledge conversion up to a certain period $t$ where $k_{i,t}=1$ if agent $i$ has become converted at time $t$ and $k_{i,t}=0$ otherwise (rules about conversion are detailed in subsequent sections). To transfer knowledge, the receiver should be able to access the knowledge. It was assumed that the information is available to the receiver when the sender meets the receiver ($m_{ij}^*=1$). That is, $p_{ij,t}$ is computed if a sender $i$ has new information and a receiver $j$ has not had the information yet, conditioned to the fact that they are expected to meet at time $t$. If one of these three condition is not satisfied at time $t$, the degree of knowledge transfer is null ($p_{ij,t}=0$).

When these conditions are met, the level of knowledge dissemination between agents is computed by the linear combination between the sender’s capability ($a_{ii}/\max_i a_{ii}$) and the extent of interaction between the sender $i$ and receiver $j$ ($a_{ij}/a_{jj}$). $\delta$ ($0<\delta<1$) is a weight for the two factors. $p_{i,t}$ is the total amount of knowledge an agent $j$ has gained from all senders $i$, such that $m_{ij}^*=1$, during time $t$.

The cumulative level of knowledge transfer up to period $t$

In this study, it was assumed that levels of disseminated knowledge can be accumulated over time. The level of the knowledge of an agent $j$ at the end of time $t$ is the sum of knowledge gained during previous periods plus newly gained knowledge during time $t$. Considering a diminishing effect of previously acquired knowledge over time, a cumulative level of knowledge of each agent $j$ up to time $t$ can be modeled as:

$$c_{j,t} = c_{j,t-1}(1 - \gamma)^{t-1} + p_{j,t}$$  \hspace{1cm} (4)

where $c_{j,t}$ is a cumulative level of knowledge gained by agent $j$ up to time $t$, where $\gamma$ is a rate of knowledge depreciation.
Agents converted up to period $t$
Once the level of cumulative knowledge reaches a certain threshold, the agent $j$ can be considered as converted, and therefore, it can start transmitting the knowledge in the next period. At the end of time $t$, a matrix $K$ is updated based on the cumulative probability $c_{j,t}$ as follows:

$$
\begin{align*}
  k_{j,t+1} &= 1 \text{ if } c_{ij,t} > \tau \\
  k_{j,t+1} &= 0 \text{ if } c_{ij,t} \leq \tau
\end{align*}
$$

(5)

where $\tau$ is a threshold for knowledge conversion. That is, $k_{j,t+1}=1$ represents that an agent $j$ is converted at the end of time $t$ (beginning of time $t+1$), and $k_{j,t+1}=0$ represents that the agent has not been converted yet by time $t$. The threshold ($\tau$) may vary, depending on the network’s tendency to accept new knowledge or concepts. Although individuals have a different level of willingness or reluctance, it can be assumed that there is an overall cut-point to determine whether an agent is converted or not.

An illustration of this procedure is shown in Figure 1.

![Figure 1: Overall methodology for updating variables.](image)

### 4 CASE STUDY AND RESULTS

#### 4.1 Case Study: Knowledge Diffusion in an Obesity Research Collaboration Network

The concepts and models presented in the methodology were tested to evaluate knowledge diffusion in a collaboration network related to obesity research in an academic medical center. For these purposes, a co-authorship-based network was generated as the baseline for understanding the current collaboration patterns. This network structure was used to approximate potential pathways of influence and communication within the network. Hence, relevant parameters to approximate intensities of collaboration, sender’s reputation, and probabilities of encounter were estimated using bibliometric information extracted
Kang and Munoz

from the co-authorship database. This database included bibliometric information for 779 articles from 1988 to 2013.

First, using the function shown in (1), the probabilities of meeting were computed (matrix M). The parameters used to model the probability of meeting were \( \varphi = 0.249 \) and \( \beta = 1.496 \). These parameters were obtained assuming that the meeting distribution passes through two points \( P_1(1,0.1) \) and \( P_2(5,0.5) \), representing that the probability of meeting during period \( t \) is 0.1 if the agents have one collaboration project and the probability of meeting during period \( t \) is 0.5 if they have five collaboration projects. For the level of knowledge transfer at a certain period \( t \) (matrix P), the weight \( \delta \) was initially set as 0.5. The impact of the parameter was tested later through a sensitivity analysis. For the cumulative level of knowledge transfer (matrix C), the knowledge depreciation rate \( \gamma \) was set at 0.05, and for agent’s conversion status (matrix K), the threshold for knowledge conversion \( \tau \) was set as 0.8.

4.2 Results

4.2.1 Collaboration Network

The obesity research collaboration network is shown in Figure 2. This network is composed of 76 researchers (nodes). The maximum number of researchers connected to a sub-network is 64. Other sub-networks connected are composed of 6, 4, and 2 researchers. The total density of the network is 0.059, which represents that approximately 5.9% of the total potential nodes currently exist within the network. The maximum geodesic distance, representing the maximum shortest path between any two nodes, is 9, which indicates that the researchers are at most separated by 8 other researchers. On average, the geodesic distance between the researchers is 3.871. The colors of the nodes represent the different clusters of the network. The Clauset-Newman-Moore algorithm was used for clustering purposes. The average size of the clusters is 8 and it ranges from 2 to 14. The average density of the clusters is 0.648 and it ranges from 0.253 to 1.

In the obesity research network graph, the size of the nodes is proportional to the number of publications of each researcher and the width of the edges are proportional to the frequency of collaboration between two researchers. The number of publications per researcher ranges from 1 to 63. Additionally, the largest intensity of collaboration between two researchers is 43. The maximum number of collaborators (degree centrality) of a researcher is 13 and the average is 4. The average number of collaborators of the top 5 researchers with more collaborators is 11. Additionally, the a group of twelve researchers were classified as experts (E) based on rules accounting for their number of publications, average number of citations of their publications, and years of experience on the field.
4.2.2 Comparison among Knowledge Dissemination Strategies

Four scenarios were proposed to reflect different knowledge dissemination strategies. These scenarios were characterized by the selection of the “seed” of knowledge, which represent researchers (agents) who are initially converted (k_{i,1} = 1) by the organization. The four proposed scenarios are: 1) top 5 agents based on degree centrality, 2) top 5 agents based on publication frequency, 3) top 5 agents based on betweenness centrality, and 4) 5 central agent based on clusters.

To evaluate the performance of the different knowledge dissemination strategies, two main performance metrics were considered: proportion of converted agents at certain periods and number of periods needed to convert a pre-specified proportion of agents. For each scenario, a simulation model was run for one year (52 weeks) with 50 replications. Table 1 summarizes key performance metrics under each of the four scenarios. The values in square brackets show 95% confidence intervals of the mean performance measures.

Table 1: Performance measure summary for each scenario.

<table>
<thead>
<tr>
<th>Key performance metric</th>
<th>Knowledge dissemination strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scenario 1</td>
</tr>
<tr>
<td>% converted at t=26</td>
<td>77.2%</td>
</tr>
<tr>
<td></td>
<td>[76.6, 77.8]</td>
</tr>
<tr>
<td>% converted at t=52</td>
<td>84.1%</td>
</tr>
<tr>
<td></td>
<td>[84.0, 84.1]</td>
</tr>
<tr>
<td>Time to convert 25%</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>[3.5, 3.7]</td>
</tr>
<tr>
<td>Time to convert 50%</td>
<td>8.6</td>
</tr>
<tr>
<td></td>
<td>[8.4, 8.8]</td>
</tr>
<tr>
<td>Time to convert 75%</td>
<td>25.9</td>
</tr>
<tr>
<td></td>
<td>[25.3, 26.5]</td>
</tr>
</tbody>
</table>

Among the scenarios, Scenario 4 had the greatest impact on disseminating knowledge within the network. After 6 months (26 weeks), about 89% of the agents were converted under scenario 4, while about 72% to 79% of agents were converted under other scenarios. After one year (52 weeks), about 92% of the agents had new knowledge under scenario 4, while 84% of the agents became converted under other
scenarios. The four scenarios showed different performances with respect to the time to convert a certain proportion of agents. At the initial dissemination phase where up to 25% agents were converted, a knowledge dissemination speed was faster under Scenarios 1 and 3 compared to Scenarios 2 and 4. However, after this initial phase, the knowledge was disseminated at a more rapid rate under Scenario 4, while it was disseminated at the slowest rate under Scenario 3. These results are supported by Figure 3 that represents the average number of researchers converted over year by scenario. Although Scenarios 1 and 2 perform better than the other two scenarios until time $t = 3$, Scenario 4 outperformed other scenarios afterwards in terms of the maximum number of converted agents as well as the speed of the dissemination. In this respect, the best strategy to disseminate knowledge within this obesity research network may be to use central agents in clusters. When initially disseminating knowledge through agents who have the highest betweenness centrality, it would take longer to spread the knowledge while relatively fewer agents would be converted. Figure 4 shows one instance of how knowledge is disseminated using scenario (strategy) 4 at periods $t = 0, 4, 8,$ and $16$ weeks.

![](image)

**Figure 3: Average number of researchers converted over a year by scenario.**

### 4.2.3 Knowledge dissemination sensitivity given different knowledge transference parameters

There are two important, but still fairly uncertain parameters in this simulation: 1) a weight ($\delta$) between a reputation factor and a collaboration intensity factor, which determine the degree of knowledge transfer, 2) a number of “seed” agents who initiate the knowledge transfer. Sensitivity analyses were conducted to understand the impact of these parameters on the key performance metrics. Figure 5 shows the average number of converted researchers by a range of the weight ($\delta$). The results indicated that the model was robust to a change in the weight between the factors affecting the extent to knowledge dissemination.
Figure 4: Instance of knowledge dissemination over time based on scenario 4.

Figure 5: Sensitivity analysis for a weight parameter (δ).

Figure 6 shows that the number of converted researchers was substantially affected by the number of initial disseminators. When fewer initiators were selected, the speed of knowledge transfer was slower and the total number of converted researchers were lower. The gap between the scenarios was significant in the beginning of the dissemination phase, but it was substantially reduced after six months. This result implies that additional research is needed to determine the optimal number of initial disseminators to achieve the best results while considering the costs and constraints based on the features of the social network in which knowledge is being disseminated.
5 CONCLUSIONS AND DISCUSSIONS

The results presented in this study provide insights into the potential impact of different knowledge dissemination strategies. This can be used at a managerial level to better utilize communications and promotional resources when disseminating knowledge (messages, practices, protocols, etc) within a network. From the case study presented, it was shown that the scenario in which the seeds of knowledge were selected based on central agents per cluster had the greatest impact on knowledge dissemination. For this scenario, the knowledge was disseminated to an average of about 66% of the network by the 8th week and to about 89% by week 26 (6 months). This kind of strategy to disseminate knowledge should be determined based on the structure of the network. In some networks, strategies of dissemination using as seeds the top agents based on degree centrality, betweenness centrality, number of publications, or other hybrid strategies could have better impact. For instance, in a fully connected network, seeds based on degree centrality or betweenness centrality might perform better as they are typically used to approximate degrees of influence and leadership of a network. The knowledge dissemination strategy can be also determined depending on other factors such as the urgency required to disseminate the knowledge, amount of resources to generate the initial seeds, priority of knowledge dissemination to certain clusters, etc. In any of these cases, the model presented can serve as the baseline for understanding and predicting the coverage and velocity of knowledge dissemination.

REFERENCES


**AUTHOR BIOGRAPHIES**

**HYOJUNG KANG,** PhD, is a postdoctoral researcher in industrial engineering and operations research at The Pennsylvania State University (PSU). She holds an M.S. in industrial and systems engineering from Georgia Institute of Technology and a B.S. in industrial and systems engineering from Yonsei University. Her research interests include agent-based modeling, discrete-event simulation, and system dynamics in healthcare. Her email address is hqk5116@psu.edu

**DAVID A. MUNOZ** (corresponding author) is a Ph.D. candidate in industrial engineering and operations research at The Pennsylvania State University (PSU). He holds an MEng. in industrial engineering from PSU and a BSc. in industrial engineering from Universidad del Bio-Bío. His research interests include
Kang and Munoz

social network analysis, dynamic network analysis, and discrete-event simulation and their applications in healthcare, manufacturing, defense, and education industries. His email address is david.munoz.soto@gmail.com