Industrial Management & Data Systems

Longitudinal analysis of online community dynamics

Seung Ik Baek Young Min Kim

The current issue and full text archive of this journal is available on Emerald Insight at:
www.emeraldinsight.com/0263-5577.htm

Longitudinal analysis of online community dynamics

Seung Ik Baek
School of Business, Hanyang University, Seoul, Republic of Korea, and
Young Min Kim
Hyundai AutoEver, Seoul, Republic of Korea

Received 20 September 2014
Revised 24 November 2014
28 January 2015
9 February 2015
Accepted 10 February 2015

Abstract

Purpose – The purpose of this paper is to explore the dynamics of an online community by examining its participants' centrality measures: degree, closeness, and the betweenness centrality. Each centrality measure shows the different roles and positions of an individual participant within an online community. To be specific, this research examines how an individual participant's role and position affects her/his information sharing activities within an online community over time. Additionally, it investigates the differences between two different online communities (a personal interest focussed community and a social interest focussed community), in terms of the interaction patterns of participants.

Design/methodology/approach – For this research, the authors collected log files from Korean online discussion communities (café.naver.com) using a crawler program. A social network analysis was used to explore the interaction patterns of participants and calculate the centrality measures of individual participants. Time series cross-sectional analysis was used to analyze the effects of the roles and the positions on their information sharing activities in a longitudinal setting.

Findings – The results of this research showed that all three centrality measures of an individual participant in previous time periods positively influenced his/her information sharing activity in the current periods. In addition, this research found that, depending on the nature of the discussion issues, the participants showed different interaction patterns. Throughout this research, the authors explored the interaction patterns of individual participants by using a network variable, the centrality, within a large online community, and found that the interaction patterns provided strong impact on their information sharing activities in the following months.

Research limitations/implications – To investigate the changes of participant's behaviors, this study simply relies on the numbers of comments received and posted without considering the contents of the comments. Future studies might need to analyze the contents of the comments exchanged between participants, as well as the social network among participants.

Practical implications – Online communities have developed to take a more active role in inviting public opinions and promoting discussion about various socio-economic issues. Governments and companies need to understand the dynamics which are created by the interactions among many participants. This study offers them a framework for analyzing the dynamics of large online communities. Furthermore, it helps them to respond to online communities in the right way and in the right time.

Social implications – Online communities do not merely function as a platform for the free exchange and sharing of personal information and knowledge, but also as a social network that exerts massive influence in various parts of society including politics, economy, and culture. Now online communities become playing an important role in our society. By examining communication or interaction behaviors of individual participants, this study tries to understand how the online communities are evolved over time.

Originality/value – In the area of online communities, many previous studies have relied on the subjective data, like participant's perception data, in a particular time by using survey or interview. However, this study explores the dynamics of online communities by analyzing the vast amount of data accumulated in online communities.

Keywords Information sharing, Online community, Social network analysis, Centrality, Time series analysis

Paper type Research paper

Industrial Management & Data Systems
Vol. 115 No. 4, 2015
pp. 661-677

0263-5577
DOI 10.1108/IMDS-09-2014-0266

1. Introduction

The internet has provided us with a variety of brand new experiences. The most notable change in our daily lives is the way people interact and communicate with each other. The internet has helped many of us communicate more effectively and economically on a two-way basis and has resulted in the formation of online communities. An online community can be defined as a group in which individuals come together around a shared purpose, interest, or goal on the internet (Preece, 2000). In the past, online communities merely served as an unofficial, passive form of media to share trivial concerns or rumors among individuals. However, they have developed to take a more active role in inviting public opinion and promoting discussion about various socio-economic issues that used to be discussed by using traditional media such as newspapers, television, and radio (Chua *et al.*, 2007; Xu and Zhang, 2009). Now online communities do not merely function as a platform for the free exchange and sharing of personal information and knowledge, but also as a social network that exerts massive influence in various parts of society including politics, economy, and culture. The influence of online communities derives from the facilitation of online communication among individuals in their daily lives, empowered by the growing use of social networking services such as Facebook or Twitter, online discussion groups such as Google or Yahoo, and a variety of message boards. Now, governments and companies begin to recognize the importance of online communities. In order to respond to online communities effectively, they need to understand online communities as complex evolving social networks in which connections between individuals are established and changed over time (Panzarasa *et al.* (2007)). In other words, they need to understand the complex social behaviors that take place in online communities from a dynamic rather than static view.

In the area of online communities, many previous studies have focussed on identifying individual's motivations to actively participate in online communities (Ardichvili *et al.*, 2003; Bateman *et al.*, 2011; Ehrlich *et al.*, 2014). Since the participation in online communities is voluntary, it is very important for community leaders or managers to understand the motivations of participants. However, the governments and the companies, who need to respond to anti-communities as well as friendly communities, are eager to know about the participants who are active in sharing information (Watts and Dodds, 2007). Because participants in online communities are physically distant and unfamiliar with each other, the social influences of the active participants are much stronger in online settings than in offline settings. However, there are few existing research works which focus on understanding the active or influential participants (Huffaker, 2010). This study, by analyzing communication behaviors and social networks of individual participants, tries to examine who will be the active or influential participants in sharing information.

As many interactions among people have been performed on the internet, the interactions have been recorded in chronological order, generating an immense amount of data. The thorough analysis of this data will help us better understand communication and social networking behaviors of participants within online communities. In order to examine the behaviors of individual participants, many previous studies have relied on subjective data, like a participant's perception data, at a particular time by using survey or interview (Hansen, 2002; Song *et al.*, 2007; Wu *et al.*, 2012). However, this study tries to explore the dynamics of online communities by analyzing the vast amount of data accumulated in online communities. In particular, by using the social network analysis (SNA), the study investigates changes in the interactive patterns among the individual participants over time. Furthermore, by using time series cross-sectional (TSCS) analysis, it tries to investigate impacts of the interaction patterns on the participant's information sharing activities in a longitudinal setting. This study focusses on the below two research questions:

RQ1. Investigating dynamics of online communities by examining interaction patterns of individual participants over time.

RQ2. Investigating impacts of interaction patterns of individual participants on their information sharing activities in the future.

2. Dynamics of online communities

In an offline-based community, the presence of physical and restricted space among participants has played an important role in building more active and vibrant communities. To form a community in an offline setting, every participant has to be close to each other in the same place at the same time. In an online community, the basic principle of physical space no longer applies, because communication on the internet goes beyond temporal and spatial limits, prompting more active interactions among a greater number of participants than in an offline setting. Lea *et al.* (2006) define an online community as "a place where a group of people drawn together by an opportunity to share a sense of community with like-mined strangers having common interest." If an online community supports interactions that provide valuable information, the community's participants are going to feel a strong sense of community, thus devoting much their time and effort in the online community (Butler, 2001; Jones, 2000). However, as many new entrants join and start making mass interactions in an online community, the quality of information might decrease. Consequently, the participants lose a sense of community and ultimately leave the online community. Under the given situation, maintaining a sense of community among participants can prove to be very difficult (Jones, 2000). In order to investigate methods to maintain a sense of community, several studies have conducted to understand the dynamics within online communities.

Jones and Rafaeli (1999), by examining the numbers of participants and messages, suggest a dynamic model of online communities. They state that, in order to initiate interactive communication within an online community, there must be a minimum number of participants, known as "Critical Mass." Normally, larger and more active online communities attract more new participants. However, beyond a certain number of participants, additional participants can create a significant adverse effect on information sharing among them. Jones and Rafaeli (1999) try to explain this phenomenon by using two factors: social loafing and information overload. Social loafing can be observed in the situations where a large majority of participants can equally enjoy benefits without offering any contributions. Social loafing in an online community is where a small number of participants tend to post a large portion of messages while a large number of participants simply read the messages without posting any messages. Another factor is information overload. Because participants have to search, filter, or process large amounts of information in a large online community, they have a tendency to experience the high communication load, resulting in fewer active participants and interactions. They find that social loafing and information overload affect the numbers of participants and messages. Butler (2001) and Jones *et al.* (2004) try to prove the effects of social loafing and information overload in online communities without considering communication behaviors of individual participants. The active participants who post many messages play more critical roles in the dynamics of online communities than other participants. After analyzing 1.03 million messages by 33,315 participants, Schoberth *et al.* (2003) find that "the 10% most active participants contribute about 70% of the message." They also find that, as time passes, the active participants dominate the online communities. By triggering message replies and spark conversation, the active participants play a crucial role in sustaining online activities (Butler *et al.*, 2002). In order to investigate the online community dynamics, it is necessary to analyze communication activities at the individual level, as well as the aggregated activity level.

Few studies explore characteristics of the active participants (Huffaker, 2010; Sudweeks and Simoff, 2005). This study considers an online community as an interpersonal social network, and tries to identify characteristics of the active participants from their roles and positions within the network to which they belong. According to Carboni and Ehrlich (2013), their roles and positions within a social network are determined by their communication patterns and affect the ability of participants to communicate with others.

3. SNA and model development

In order to examine the interactions among online community participants and their changes, this study uses SNA to examine a large amount of user-generated online messages. SNA was first conceived and developed in the 1930s by social science scholars, it but did not receive much scholarly attention until the late twentieth century when numerous notions, such as Sociometry, Graph Theory, Dyads, Triads, Subgroup, and Blockmodels, were theorized. Currently, SNA is applied to a wide range of studies including sociology, business administration, and economics (Kane and Alavi, 2005; Mutuswami and Winter, 2002). The primary concern of SNA-related research has been the "interpersonal connections" through "interactions." Researchers have attempted to quantitatively analyze the structure of such connections, mostly using Graph Theory that describes the links (interactions) among the nodes (participants) within a network. This theory enables researchers to explain the structure of social networks by transforming them into numerical indicators showing the various graphical features displayed by the push and pull between the nodes and the links (Wasserman and Faust, 1994). SNA helps us to have a better understand of participants' interactions and information sharing activities (Sitko-Lutek *et al.*, 2010). Table I is a summary of the social network indicators at the level of individual and network.

This study tries to explore the dynamics of online communities by examining interaction activities of participants. In order to investigate the interaction activities, it

Unit of analysis	Network indicators	Explanations
Individuals	Centrality	Power/popularity of individual participants
Networks	Density	Frequency of interactions among the participants within a community
	Centralization	Degree of centralization of an individual participant within a community
	Reciprocity	Degree of mutual interactions among participants within a community
	Core-periphery	Number of sub-groups within a community

Table I.
Typology of social network measures

utilizes an individual-level network indicator, centrality. Centrality indicates how well positioned an individual participant is to retrieve and disseminate information within a community (Wasserman and Faust, 1994). A large body of research has shown that people who are centrally located in a network can access more information easily, leading to more opportunities and advantages in performing tasks than people who are peripheral in the networks (Ahuja *et al.*, 2003; Chung and Hossain, 2009; Maria, 2010). Depending on the communicative positions in a network, Freeman (1979) identified three different centrality measures: degree centrality, closeness centrality, and betweenness centrality.

(1) Degree centrality

Degree centrality is an indicator that accounts for the number of direct links an individual participant has with others, thus indicating the influence of a certain individual in a network. In other words, degree centrality implies how active a particular participant is. An individual with a high degree centrality might be the leader or the hub in his/her group or network. Since he/she has many direct links with other members, he/she can easily access more information and be reached by others easily. There are two types of degree centrality depending on the direction of the links: out-degree centrality and in-degree centrality. If an individual participant initiates his/her interactions by posting messages to many other participants, he/she has a higher out-degree centrality. On the other hand, if an individual participant receives many messages from other participants because of his/her popularity and knowledge, he/she has a higher in-degree centrality. While a participant's in-degree centrality is a good indication of his/her popularity in a network and accessibility to information, a participant's out-degree centrality is a good indication of his/her control over a network and the dependence of the network upon him/her (Loosemore, 1998). Based on the formula shown below, we can calculate the degree centrality of participants:

$$C_D(P_k) = \frac{\sum_{i=1}^{n} (P_i, P_k)}{(n-1)}$$

where $C_D(P_k)$ is the degree centrality of P_k; P_k is a participant; $(P_i, P_k) = 1$ (if a link exists between P_i and P_k); and $(P_i, P_k) = 0$ (if a link does not exists between P_i and P_k).

Since the participants who have high in-degree or out-degree centrality are in direct contacts with many other participants within the networks, the participants are immediately recognized by others as a hub, increasing their status, reputation, and ability to influence others (Butler *et al.*, 2002; Huffaker, 2010). Preferential Attachment Theory states that "participants with more existing links have a higher probability of receiving additional links than participants with fewer existing links" (Lu *et al.*, 2013):

H1. In-degree centrality of a participant in previous time periods provides a positive impact on the number of his/her messages made in the following time periods.

H2. In-degree centrality of a participant in previous time periods provides a positive impact on the number of his/her messages received in the following time periods.

H3. Out-degree centrality of a participant in previous time periods provides a positive impact on the number of his/her messages made in the following time periods.

H4. Out-degree centrality of a participant in previous time periods provides a positive impact on the number of his/her messages received in the following time periods.

(2) Closeness centrality

Closeness centrality is the sum of the length of the geodesics between a particular node and all the other nodes in a network. It measures how far away one node is from the other nodes and is sometimes called "farness" (Freeman, 1979). Higher closeness implies shorter, less expensive, and more efficient paths in receiving or providing information (Song *et al.*, 2007). It focusses on the extent of influence over the entire network (Yang and Ding, 2009). The participants with higher closeness are likely to directly monitor and control a great number of other participants, and to quickly disseminate decisions and ideas to a wider range of participants. Depending on the direction of the links, there are two types of closeness centrality: out-closeness centrality and in-closeness centrality. While a participant with high in-closeness centrality might listen to many other participants through either direct or indirect links, a participant with high out-closeness centrality might post messages to many other participants through either direct or indirect links. Wasserman and Faust (1994) argue that, while the degree view of centrality represents being influential and respectable, the closeness view focusses on the economic considerations of communication. Based on the below formula, we can calculate the closeness centrality of participants:

$$C_c(P_k) = \frac{1}{\sum_{i=1}^{n} d(P_i, P_k)}$$

where $C_C(P_k)$ is closeness centrality of P_k; P_k is a participant; and $d(P_i, P_k)$ is length of the shortest path between P_i, and P_k.By definition, the participants who have high closeness centrality can quickly interact with participants who are not first neighbors (Latora and Marchiori, 2008). In other words, they can reach all the other participants in as few steps as possible. Like the degree centrality, the closeness centrality allows them to access more information easily. In addition, because the participants with high closeness centrality have many direct and indirect connections with others, they can exert influence on more participants than the participant with high degree centrality (Dogan *et al.*, 2013). However, Ibarra and Andrews (1993) insist that the high closeness always does not guarantee strong power. Although participants have low closeness centrality, if they have the closest relations (direct or short indirect links) with many central participants, they can have strong power in a network. Therefore, this study hypothesizes that:

H5. In-closeness centrality of a participant in previous time periods provides a positive impact on the number of his/her messages made in the following time periods.

H6. In-closeness centrality of a participant in previous time periods provides a positive impact on the number of his/her messages received in the following time periods.

H7. Out-closeness centrality of a participant in previous time periods provides a positive impact on the number of his/her messages made in the following time periods.

H8. Out-closeness centrality of a participant in previous time periods provides a positive impact on the number of his/her messages received in the following time periods.

(3) Betweenness centrality

Betweenness centrality measures the extent to which a particular node lies between the other nodes in a network or a group. The betweenness centrality of a node is defined as the number of geodesics (shortest paths between two nodes) passing through it. In other words, this score represents the number of times that a participant needs another given actor to reach any other participant by the shortest path. A participant with higher betweenness centrality may act as a gatekeeper or broker in a network. According to Freeman (1979), the participants with higher betweenness are likely to have the potential to control others by controlling and filtering information flows within a network. Based on the formula below, we can calculate the betweenness centrality of participants:

$$C_B(P_k) = \frac{\left(\sum_{s<t} n_{st}(P_k)\right)/n_{st}}{((n-1)(n-2))/2}$$

where $C_B(P_k)$ is betweenness centrality of P_k; P_k is a participant; $n_{st}(P_k)$ is number of the shortest paths between node s and node t ($s \neq t$) passing through P_k; n_{st} is number of all shortest paths between node s and node t; and $(n-1)(n-2)/2$ is maximum betweenness.

People who have high betweenness centrality are located at the intersections between two non-adjacent points, allowing for the control of information between two points (Latora and Marchiori, 2008). Unlike degree centrality and closeness centrality, the betweenness centrality provides participants with diverse resources located in multiple sub-groups (Cho *et al.*, 2007). Having the ability to access weakly connected sub-groups create a powerful and influential position for the brokers. Although they are not at the center of the networks, they play a crucial role in disseminating and regulating information at their will. Park and Suh (2013) provide an empirical evidence that the betweenness centrality can be used for indicating influential participants in an online social networking service. In addition, Abbasi *et al.* (2012) find that, by examining structural changes in a scientific co-authorship network, new entrants prefer to attach to the participants who have higher betweenness centrality rather than those with higher degree or closeness centrality. Therefore this study hypothesizes that:

H9. Betweenness centrality of a participant in previous time periods provides a positive impact on the number of his/her messages made in the following time periods.

H10. Betweenness centrality of a participant in previous time periods provides a positive impact on the number of his/her messages received in the following time periods.

As the participants establish new relations with others by posting messages, their social positions, roles, and power may change accordingly. These node dynamics resulting from relation changes can be captured by three centrality measures.

4. Research methodology

The communication structure within online communities keeps evolving, driven by the constant interactions among the participants. Online communications in the communities primarily take the form of commenting on an internet bulletin board. An analysis of the commenting process offers a better grasp of the interactive characteristics of the community participants. Therefore, this study looks closely into the commenting behaviors among the online community participants, exploring how their behaviors change over time.

In order to investigate how the different issues of online communities affect the participants' information sharing behaviors, this study chose two online communities from a popular Korean popular portal site (www.naver.com). As a personal interest focussed community, we chose an online community for digital cameras. In the community, participants shared various information about new digital cameras or shooting methods. As a social issue focussed community, we chose an online community in which participants discussed public lies about one of the Korean pop artists.

The discussion boards in Naver are organized in the following structure: a participant raises an issue by posting on the bulletin board, triggering others to interact by replying to each other's writings on the topic. Figure 1 illustrates how an initial post-branches out into a series of messages. If a participant starts a thread and receives three direct replies (Replies 1-3), this study assumes that he/she make direct interactions with three reply posters. Furthermore, because Replies 1a and 1b are associated with an initial thread through Reply 1, a participant who initiates the discussion interacts with posters of Replies 1a and 1b indirectly. The interactive patterns among participants are analyzed through the SNA method by scrutinizing the structure of the replies. In order to use the SNA method, at first we need to construct an adjacency matrix, in which both the columns and rows represents participants and a "1" in the cells represents the presence of interaction between the two participants (refer to Figure 2). With the help of the adjacency matrix, three centrality variables – degree, closeness, and between centrality – are calculated using NetMiner. Finally we construct adjacency matrixes among the 112,616 participants from the personal interest focussed community and among the 102,452 participants from the social issue focussed community. Table II summarizes the data collection strategy for this study. In order to investigate the changes in interaction among the participants over time, we collected an activity log data of the participants over the course of 15 months.

For this study, we collected log files from two online communities operated by a Korean internet portal site, Naver (café.naver.com) (refer to Table II). Naver, one of the top Korean internet portal sites, currently operates 10,000,000 communities in various areas. In some of these communities, people have shared their own opinions against government policies or companies beyond personal interests. Online communities in Naver are the most active communities in Korea.

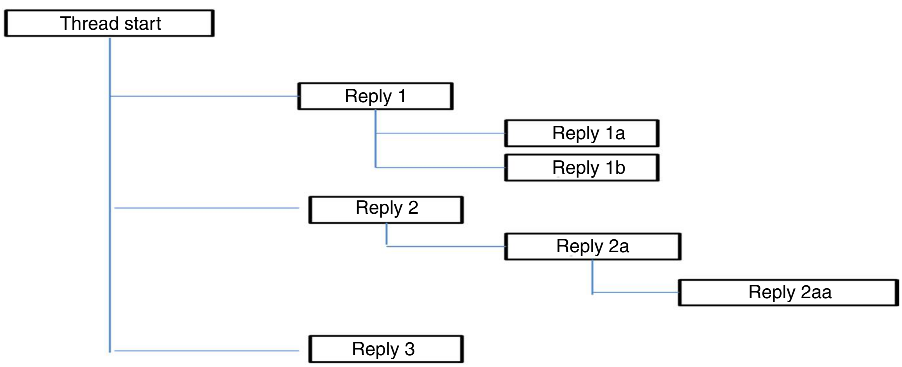

Figure 1.
Interaction structures in discussion board

Figure 2.
Adjacency matrix

Table II.
Data collection strategy

Type	Community	Period	Data sources	Data collected
Personal issue	Cannon DSLR Club	15 months	http://cafe.naver.com/canon450dclub	From discussion board IDs of participants
Social issue	Public Lies about a Celebrity	15 months	http://cafe.naver.com/tajinyo3	No of participants No of issues and replies

A SNA was used to calculate the centrality measures of the individual participants, and TSCS analysis was used to analyze the effects of centralities of participants on their future information sharing activities.

5. Data analysis

(1) Number of participants

Figure 3 shows the number of participants in two online communities. While the number of participants in the personal issue focussed community was been very stable over 15 months, the number of participants in the social issue focussed community fluctuated. In the social issue focussed community, many participants joined

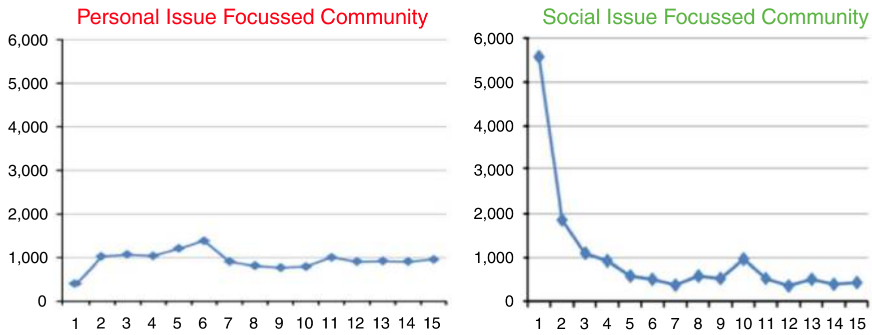

Figure 3.
Number of participants

the community from the beginning, but lost their interests in the issue rapidly as the discussions continued. Many participants shared information with others most actively during the first two months.

(2) Network structure analysis

Based on the adjacency matrixes, 15 network diagrams (one network diagram for each month) were created for each community. Figure 4 shows the changes in the interaction patterns over 15 months. Within a social issue focussed community, all the participants actively interact with others from the beginning. However, as time passes, the interactions between the participants become weak. In contrast, within a personal interest focussed community, as time passes, the interactions become strong. The darker areas in the network diagrams indicate intensity of interactions among participants. Both communities have one similarity: several opinion leaders or hubs have appeared as the discussion has progressed.

(3) Correlation analysis

Before conducting a time series analysis, we investigate the multicollinearity problems by analyzing the correlations among the variables. Tables III and IV show the results of

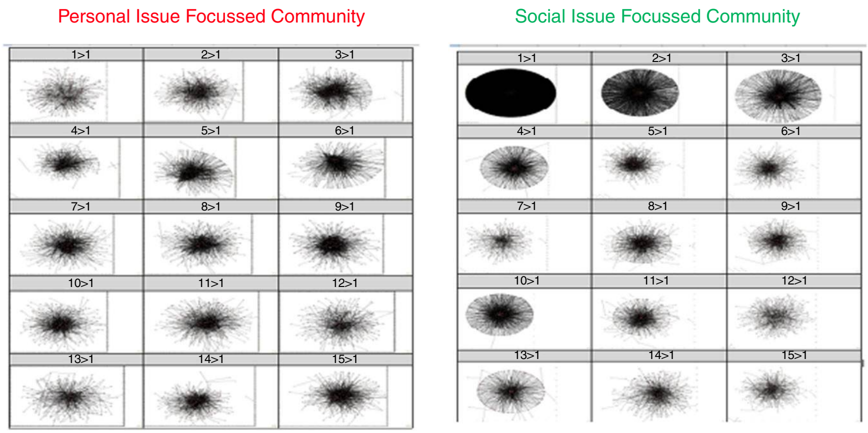

Figure 4.
Network diagrams

Table III.
Correlation matrix of a personal interest focussed community

	(1) No. of messages received	(2) No. of messages posted	(3) In-degree centrality	(4) Out-degree centrality	(5) In-closeness centrality	(6) Out-closeness centrality	(7) Betweenness centrality
(1)	1.00						
(2)	0.495**	1.00					
(3)	0.548**	0.248**	1.00				
(4)	0.285**	0.418**	0.431**	1.00			
(5)	0.324**	0.188**	0.711**	0.322**	1.00		
(6)	0.251**	0.302**	0.409**	0.777**	0.376**	1.00	
(7)	0.358**	0.395**	0.549**	0.804**	0.297**	0.545**	1.00

Note: $^{**}t < 0.01$

the correlation analysis. As shown in the tables, all five centrality measures were significantly correlated with two information sharing activities in both communities. In particular, in-degree centrality had the strongest correlation with the number of messages being received, and out-degree centrality had the strongest correlation with the number of messages being posted. This means that the participants with many direct links to other participants play a hub by leading the discussions within the online communities. To check the multicollinearity problems among the independent variables, this study uses variance inflation factor (VIF). Tables V and VI show VIFs of the two communities. Since all VIFs are less than ten, the independent variables of the study appear not to have multicollinearity problems.

(4) Time series analysis

To examine the hypotheses that the centrality measures (positions and roles) of each participant in previous time periods $(t-1)$ do affect his/her information sharing activities in the current time periods (t), a time series analysis was conducted. The data

	(1) No. of messages received	(2) No. of messages posted	(3) In-degree centrality	(4) Out-degree centrality	(5) In-closeness centrality	(6) Out-closeness centrality	(7) Betweenness centrality
(1)	1.00						
(2)	0.648**	1.00					
(3)	0.520**	0.281**	1.00				
(4)	0.309**	0.310**	0.517**	1.00			
(5)	0.268**	0.131**	0.626**	0.343**	1.00		
(6)	0.271**	0.210**	0.514**	0.648**	0.522**	1.00	
(7)	0.384**	0.310**	0.698**	0.778**	0.319**	0.426**	1.00

Note: $^{**}t < 0.01$

Table IV. Correlation matrix of a social issue focussed community

Independent variables	VIF
Intercept	0
In-degree centrality	2.61154
Out-degree centrality	5.17491
In-closeness centrality	2.07523
Out-closeness centrality	2.79093
Betweenness centrality	3.53871

Table V. VIFs of a personal interest focussed community

Independent variables	VIF
Intercept	0
In-degree centrality	3.52052
Out-degree centrality	4.29729
In-closeness centrality	2.06468
Out-closeness centrality	2.48279
Betweenness centrality	4.59539

Table VI. VIFs of a social issue focussed community set of this study consists of time series observations (three centrality measures and the numbers of messages received/posted by each participant over 15 months) on each of the cross-sectional units (individual participants). This type of data is called the panel data. The data includes multiple entities, each of which has repeated measurements at different times. The panel data set has both a cross-section variable and a time-series variable. This study uses online community participants as cross-section variables and one month as a time-series variable. Given the structure of our data, we use a panel data analysis technique, the time series cross-section regression (TSCSREG) procedure in SAS (Parks, 1967; SAS, 1993). Table VII summarizes the results of TSCSREG. According to the results, in-degree centrality, out-degree centrality, and betweenness centrality provide a statistically significant impact on the numbers of messages received and messages posted in both communities. In particular, in-degree centrality offers the strongest impact on the number of received messages. In terms of the number of posted messages, out-degree centrality offers the strong impact (in the social issue focussed community, there is a relatively small difference between in-degree centrality and out-degree centrality). The existing studies showed that, because the participants who have high degree centrality (in-degree and out-degree) have direct links with many participants, they are the hub of communication activity (Ahuja *et al.*, 2003; Song *et al.*, 2007). On the other hand, in-closeness centrality and out-closeness centrality show mixed results (some are negative relationships and some are non-significant relationships). For the unexpected results from previous studies, we have a possible explanation. The existing studies investigated the relationship between closeness centrality and work performance in offline settings, not in online settings (Hansen, 2002; Song *et al.*, 2007). However, this study analyzes network structures from the discussion boards of online communities. Since the discussion boards show messages in the order of posting, there is a possibility that participants responded to the recent postings without considering the influence of message contributors or message contents. While the effect of direct links (degree centrality) is likely to be overestimated, one of indirect links (closeness centrality) is likely to be underestimated. Based on the result, we find that the effect of local centrality (degree centrality) is much stronger than one of global centrality (closeness centrality) in discussion board-based communities. If this study is conducted with Twitter or Facebook, in which messages are organized by contributors as well as topics, it might have different results.

In order to look closely at individual relationships between each centrality measure and information sharing activities, we conducted a time series analysis for each centrality measure one by one. Tables VIII-X summarize the results of the time series analyses. The results show that all three centrality measures of a participant in

		Independent variables ($t-1$ period)					
		IN-D	OUT-D	IN-C	OUT-C	BET	R^2
Personal interest	No. of messages received	65.36**	2.4**	−12.45**	3.91**	16.6**	0.475
Focussed comm.	No. of messages posted	8.3**	24.98**	6.74**	−6.7**	19.67**	0.305
Social issue	No. of messages received	69.03**	11.92**	−6.89**	−0.49	9.35**	0.433
Focussed comm.	No. of messages posted	15.99**	14.08**	4.48**	−1.58	6.4**	0.256

Notes: IN-D, in-degree centrality; OUT-D, out-degree centrality; IN-C, in-closeness centrality; OUT-C, out-closeness centrality; BET, betweenness centrality. **$t < 0.01$

Table VII.
Results of TSCSREG

previous time periods ($t-1$) provide a positive impact on the numbers of messages posted and received during current time periods (t). They even show the significant impact of closeness centrality. In terms of the effects of three centrality measures on the number of messages received and received between the two communities, they do not show any significant differences.

6. Conclusions and implications

This study tries to understand the dynamics from a social network perspective. In particular, it focusses on identifying who lead and show one creates the dynamics. The study finds that, as the discussion progresses, interactions among participants rapidly change. In addition, it demonstrates that participants actively interact with others in the beginning, but most interactions converge among a few active participants eventually. Interestingly, it shows that certain participants emerge to lead the dynamics of online communities by actively participating in information sharing.

			Independent variables ($t-1$ period)	
			In-degree centrality	Out-degree centrality
Dep. variables (t period)	Personal interest	No. of messages received	94.52**	21.95**
		No. of messages posted	26.44**	63.34**
	Social issue	No. of messages received	88.48**	9.91**
		No. of messages posted	29.51**	28.98**

Note: $^{**}t < 0.01$

Table VIII. Relationships between degree centrality and information sharing activity

			Independent variables ($t-1$ period)	
			In-closeness centrality	Out-closeness centrality
Dep. variables (t period)	Personal interest	No. of messages received	49.64**	35.15**
		No. of messages posted	22.17**	44.54**
	Social issue	No. of messages received	33.65**	28.48**
		No. of messages posted	16.78**	20.53**

Notes: $^{**}t < 0.01$

Table IX. Relationships between closeness centrality and information sharing activity

			Independent variables ($t-1$ period)
			Betweenness centraility
Dep. variables (t period)	Personal interest	No. of messages received	84.2**
		No. of messages posted	81.77**
	Social issue	No. of messages received	65.6**
		No. of messages posted	48.98**

Note: $^{**}t < 0.01$

Table X. Relationships between betweenness centrality and information sharing activity

This study offers three theoretical implications for researchers. First, it provides a new perspective in understanding the dynamics of online communities. Existing studies have tried to understand a dynamic model of online community by examining communication activities at an aggregate level, not an individual level (Butler, 2001; Jones and Rafaeli, 1999; Schoberth *et al.*, 2003). By aggregating individual communication activities, they explored various dynamic changes in online communities (e.g. the numbers of participants, active participants, new participants, and messages). Unlike the existing studies, this study focusses on analyzing individual participant's activities. Second, this study considers a time factor to examine dynamics of online communities. Previous studies in the area of SNA have explored the relationship between participant's centrality and his/her work performance without considering a time factor (Carboni and Ehrlich, 2013; Song *et al.*, 2007; Wu *et al.*, 2012). Previous studies have validated the existence of strong links between centrality measures and work performance in various settings. However, previous studies have ignored the fact that the network structures in online communities change rapidly over time. The third theoretical implication is in the area of leadership in online communities. Sudweeks and Simoff (2005) categorized online leadership into two types: assigned leadership and emergent leadership. Sudweeks and Simoff (2005) defined an emergent leader as "an individual who is not assigned to a leadership position initially, but gradually emerges as a leader throughout the support and acceptance of other participants." Lu *et al.* (2013) and Yoo and Alavi (2004), by examining communication activities of the emergent leaders, tried to distinguish emergent leaders from non-leaders. This study tries to suggest a predicator of emergent leaders by using SNA.

This study provides some practical implications for companies or governments managing online communities. As many people continue to use online communities as major channels to share their opinions and interact with others, companies, and governments can no longer ignore the influence of online communities. Online communities have played critical roles in forming public opinions. As various social media services are supporting activities in online communities, public opinions from online communities can threaten the survival of companies and governments. Given the situation, companies and governments may need to monitor the activities of online communities closely. However, they lack any guidelines in how to respond to online communities (Elberse, 2008). In addition, they need to identify active or influential participants who post many messages and receive attention from other participants. Recently, many research has been trying to propose methods of identifying possible influential participants (Lu *et al.*, 2013; Sung *et al.*, 2013). This study proposes three network measures as predicators to distinguish influential participants.

7. Limitations

This study has two limitations. First, it focusses on examining two online communities on a Korean portal site. Although the research site is one of the largest and the most popular sites in Korea, the study still has the problem of generalizing the findings of this research. Cultural factors could possibly affect the attitude of participants in sharing their opinions on the internet (Jiacheng *et al.*, 2010). Therefore, we can generalize the findings of this study in a very limited way and scope.

The second limitation is closely related with the problem of SNA as a research methodology. SNA is considered a set of powerful techniques among researchers who are interested in various online related issues, because it allows them to harness a large amount of data without having to collect extra data (Howison *et al.*, 2012). Since many

SNA based studies in IS areas normally use the system-generated log data, it is very difficult to verify its quality and reliability. To investigate the changes of participant's behaviors, this study simply relies on the numbers of messages received and posted without considering the contents of the messages. Future studies might need to analyze the contents of the messages exchanged between participants, as well as the social network among participants. In addition, there is high possibility that personal perceptions and beliefs affect information sharing behaviors (Bateman *et al.*, 2011; Ehrlich *et al.*, 2014). If we could include the participants' personal information with the SNA, we could get a more comprehensive view of online community dynamics.

References

Abbasi, A., Hossain, L. and Leydesdorff, L. (2012), "Betweenness centrality as a driver of preferential attachment in the evolution of research collaboration networks", *Journal of Informetrics*, Vol. 6 No. 3, pp. 403-412.

Ahuja, M.K., Galletta, D.F. and Carley, K.M. (2003), "Individual centrality and performance in virtual R&D groups: an empirical study", *Management Science*, Vol. 49 No. 1, pp. 21-38.

Ardichvili, A., Page, V. and Wentling, T. (2003), "Motivation and barriers to participate in virtual knowledge-sharing communities of practice", *Journal of Knowledge Management*, Vol. 7 No. 1, pp. 64-77.

Bateman, P.J., Gray, P.H. and Butler, B.S. (2011), "Research note-the impact of community commitment on participation in online communities", *Information Systems Research*, Vol. 22 No. 4, pp. 841-854.

Butler, B. (2001), "Membership size, communication activity, and sustainability: a resource-based model of online social structures", *Information Systems Research*, Vol. 12 No. 4, pp. 346-362.

Butler, B., Sproull, L., Kiesler, S. and Kraut, R. (2002), "Community effort in online groups: who does the work and why?", in Weisband, S. (Ed.), *Leadership at a Distance: Research in Technologically Supported Work*, pp. 171-194.

Carboni, I. and Ehrlich, K. (2013), "The effect of relational and team characteristics on individual performance: a social network perspective", *Human Resource Management*, Vol. 52 No. 4, pp. 511-535.

Cho, H., Gay, G., Davidson, B. and Ingraffea, A. (2007), "Social networks, communication styles, and learning performance in a CSCL community", *Computers & Education*, Vol. 49 No. 2, pp. 309-329.

Chua, Z., Goh, K.Y., Kankanhalli, A., and Phang, C.W. (2007), "Investigating participation in online policy discussion forums over time: does network structure matter", *Proceedings of International Conference on Information Systems*, Montreal, pp. 117-135.

Chung, K.S.K. and Hossain, L. (2009), "Measuring performance of knowledge-intensive workgroups through social networks", *Project Management Journal*, Vol. 40 No. 2, pp. 34-58.

Dogan, S.Z., Arditi, D., Gunhan, S. and Erbasaranoglu, B. (2013), "Assessing coordination performance based on centrality in an e-mail communication network", *Journal of Management in Engineering*, Vol. 10, available at: http://dx.doi.org/10.1061/(ASCE)ME.1943-5479.0000255

Ehrlich, K., Muller, M., Matthews, T., Guy, I. and Ronen, I. (2014), "What motivates members to contribute to enterprise online communities?", *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pp. 149-152.

Elberse, A. (2008), "Should you invest in the long tail?", *Harvard Business Review*, Vol. 86, July-August, pp. 88-96.

Freeman, L.C. (1979), "Centrality in social networks-conceptual clarification", *Social Networks*, Vol. 1 No. 3, pp. 215-239.

Hansen, M.T. (2002), "Knowledge networks: explaining effective knowledge sharing in multiunit companies", *Organization Science*, Vol. 13 No. 3, pp. 232-248.

Howison, J., Wiggins, A. and Crowston, K. (2012), "Validity issues in the use of social network analysis for the study of online communities", *Journal of the Association of Information Systems*, Vol. 12 No. 12, pp. 767-797.

Huffaker, D. (2010), "Dimensions of leadership and social influence in online communities", *Human Communication Research*, Vol. 36 No. 4, pp. 593-617.

Ibarra, H. and Andrews, S.B. (1993), "Power, social influence, and sense making: effects of network centrality and proximity on employee perceptions", *Administrative Science Quarterly*, Vol. 38 No. 2, pp. 277-303.

Jiacheng, W., Lu, L. and Francesco, C. (2010), "A cognitive model of intra-organizational knowledge-sharing motivations in the view of cross-culture", *International Journal of Information Management*, Vol. 30 No. 3, pp. 220-230.

Jones, Q. (2000), "Time to split, virtually: expanding virtual publics into vibrant virtual metropolises", *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*, IEEE, Hawaii, pp. 1-10.

Jones, Q. and Rafaeli, S. (1999), "User population and user contributions to virtual publics: a systems model", *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work*, ACM, Phoenix, AZ, pp. 239-248.

Jones, Q., Ravid, G. and Rafaeli, S. (2004), "Information overload and the message dynamics of online interaction spaces: a theoretical model and empirical exploration", *Information Systems Research*, Vol. 15 No. 2, pp. 194-210.

Kane, G. and Alavi, M. (2005), "Casting the net: a multimodal network perspective on knowledge management", *Proceedings of International Conference on Information Systems*, Las Vegas, NV, pp. 233-246.

Latora, V. and Marchiori, M. (2008), "A measure of centrality based on the network efficiency", *New Journal of Physics*, Vol. 9 No. 6, pp. 1-16.

Lea, B., Yu, W. and Maguluru, N. (2006), "Enhancing business networks using social network based virtual communities", *Industrial Management & Data Systems*, Vol. 106 No. 1, pp. 121-138.

Loosemore, M. (1998), "Social network analysis: using quantitative tool within an interpretative context to explore the management of construction crises", *Engineering, Construction and Architectural Management*, Vol. 5 No. 4, pp. 315-326.

Lu, Y., Jerath, K. and Singth, P. (2013), "The emergence of opinion leaders in a networked online community: a dyadic model with time dynamics and a heuristic for fast estimation", *Management Science*, Vol. 59 No. 8, pp. 1783-18799.

Maria, E.E. (2010), "Is centrality the key to high performance", *Journal of Competitiveness*, Vol. 2, pp. 58-73.

Mutuswami, S. and Winter, E. (2002), "Subscription mechanisms for network formation", *Journal of Economic Theory*, Vol. 106 No. 2, pp. 242-264.

Panzarasa, O., Opsahl, T. and Carley, K.M. (2007), "Patterns and dynamics of users' behavior and interaction: network analysis of an online community", *Journal of the American Society for Information Science and Technology*, Vol. 60 No. 5, pp. 911-932.

Parks, R.W. (1967), "Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated", *Journal of the American Statistical Association*, Vol. 62 No. 318, pp. 500-509.

Park, J.H. and Suh, B. (2013), "The impact of influential's betweenness centrality on the WOM effect under the online social networking service environment", *Proceedings of Pacific Asia Conference on Information Systems*.

Preece, J. (2000), *Online Communities: Designing Usability, Supporting Sociability*, John Wiley and Sons, Chichester.

SAS (1993), *User's Guide, Version 6*, 2nd ed., SAS Institute, Cary, NC.

Schoberth, T., Preece, J. and Heinzl, A. (2003), "Online communities: a longitudinal analysis of communication activities", *Proceeding of the 36th Hawaii International Conference on System Sciences*, Big Island, Hawaii, pp. 1-10.

Sitko-Lutek, A., Chuancharoen, S., Sukpitikul, A. and Phusavat, K. (2010), "Applying social network analysis on customer complain handling", *Industrial Management and Data Systems*, Vol. 110 No. 9, pp. 1402-1419.

Song, S., Nerur, S. and Teng, J.T. (2007), "An exploratory study on the roles of network structure and knowledge processing orientation in work unit knowledge management", *ACM SIGMIS Database*, Vol. 38 No. 2, pp. 8-26.

Sudweeks, F. and Simoff, S.J. (2005), "Leading conversations: communication behaviours of emergent leaders in virtual teams", *Proceedings of the 38th Annual Hawaii International Conference on System Science*, Big Island, Hawaii, pp. 1-10.

Sung, J., Lee, J. and Lee, U. (2013), "Booming up the long tails: discovering potentially contributive users in community-based question answering services", *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, pp. 602-610.

Wasserman, S. and Faust, K. (1994), *Social Network Analysis Methods and Applications*, Cambridge University Press, New York, NY.

Watts, D. and Dodds, P. (2007), "Influentials, networks, and public opinion formation", *Journal of Consumer Research*, Vol. 34 No. 4, pp. 441-458.

Wu, W., Yeh, R. and Hung, H. (2012), "Knowledge sharing and work performance: a network perspective", *Social Behavior and Personality*, Vol. 40 No. 7, pp. 1113-1120.

Xu, S.X. and Zhang, X. (2009), "How do social media shape the information environment in the financial market?", *Proceedings of International Conference of Information Systems*, Phoenix, AZ, pp. 1-16.

Yang, E. and Ding, Y. (2009), "Applying centrality measures to impact analysis: a coauthorship network analysis", *Journal of the Association for Information Science and Technology*, Vol. 60 No. 10, pp. 2107-2118.

Yoo, Y. and Alavi, M. (2004), "Emergent leadership in virtual teams: what do emergent leaders do?", *Information and Organization*, Vol. 14 No. 1, pp. 27-58.

Corresponding author
Professor Seung Ik Baek can be contacted at: sbaek@hanyang.ac.kr