Discovering and Analyzing Methods on User-innovation Knowledge in Virtual Communities Based on Weighted Knowledge Network

Xiao Liao, Zhihong Li and Yunjiang Xi*

School of Business Administration, South China University of Technology, Guangzhou, 510641, China

Abstract: Considering the importance of user-innovation knowledge to product and service innovation, this paper proposed knowledge discovering and analyzing methods on user-innovation communities based on weighted knowledge network (WKN). First, web mining methods were used for getting forum posts in virtual user-innovation communities, then the acquired keywords and the statistical data were processed and integrated into a WKN model. Next, based on the WKN model, the paper proposed new methods to discover the core innovation knowledge, the developing innovation knowledge and the potential innovation knowledge through analyzing the structure of WKN model, including weights of nodes and edges, cliques, the strong or weak ties, etc. Finally, Xiaomi community was taken as the study case to verify the validity of the methods. Compared with the existing method of user knowledge discovery, the presented methods are more quantitatively, conveniently and visually and provide new tools to study and utilize the user-innovation knowledge in enterprise virtual communities.

1. INTRODUCTION

User innovation is being regarded as the important strategy in company's product and service innovation. With the spread of internet, companies are becoming aware of utilizing internet user communities to identify user innovations. Füller *et al.* [1] indicated that the internet offers new simplified modes of interaction between producers and consumers in a large scale and introduced the community based innovation (CBI) method after empirical research on firms such as Audi and BMW. According to this method, in online communities, there exist innovative customers who are not only knowledgeable but also able to develop or innovate their own new products. Thus, companies need to identify and access these online communities and interact with the community members so as to get valuable inputs for their new product development (NPD).

Nowadays, companies are beginning to offer users their own communities on their company websites which are more convenient to interact with users and get valuable ideas or thoughts for product innovations or new product development. Such enterprise virtual communities are based upon users or other people who are interested in the products. They exchange knowledge about products using experiences, share solutions to product using problems, and ideas of modification for the products or even entirely new product designs in enterprise virtual communities, then companies can utilize them in their product innovation or NPD process. Although much work has been undertaken on the development

of user innovation such as providing a number of tools to interact with users and activate the involvement of users to the company's product innovation, little research has been done on identifying and integrating the user-innovation knowledge from enterprise virtual communities. Thus, this paper aims to present a comprehensive method of mining and integrating user-innovation knowledge which combines the approaches of Web content mining, keyword analysis, weighted knowledge network and social network analysis and finally be verified with the help of a case study on Xiaomi community in China.

2. LITERATURE REVIEW

2.1 User Innovation

A great deal of research has confirmed that users play an important role in the company's innovation process. It has been found that innovations developed by users are often of high quality [2]. von Hippel [3] found that users were the developers of 67% of the major innovations in semiconductor processing and also the developers of about 80% of the most important scientific instrument innovations. Shah [4] also found that users were innovators on the most commercially important equipment in 4 sporting fields.

Now, in many fields, more and more product users engage in modifying or developing entirely new products they used. Researchers have provided lots of evidence in support of users' innovative potential [2-11]. Here, user was defined by von Hippel as firms or individual consumers that expect to benefit from using a design, a product or a service [8]. Among the innovative users, there are lead users who serve as a need-forecasting laboratory for marketing research and

can provide new product concept and design as well [8]. Therefore, many researches were undertaken to identify lead users and user needs *via* the lead user theory [3, 11, 12].

2.2 Community Based Innovation

With the spread of internet, users with similar interests in improving products or developing innovations in a specific area join together to some innovation communities where they can share what they have developed or innovated to the community, or conduct innovation projects collaboratively. Franke and Shah [13] found that on average one third of the members in those innovation communities improve their sports equipment or even design their own products. Füller *et al.* [14] proposed the community based innovation (CBI) theory which was a method to make good use of the innovative potential of virtual communities. According to this theory, companies need to identify and get access to the communities and interact with the members to get valuable inputs.

Kozinets, R. [15] introduced the symbolism, meanings, and consumption patterns of online communities and described methods to identify and get in contact with online communities. Kuo-Ming Chu &Hui-Chun Chan [16] designed a research model to examine the antecedents of CBI (community based innovation) in online communities, and analyzed how firms can best interact with online communities to get valuable inputs for their new product development, and finally found that besides the individual differences of communities members and the characteristics of products or industries, the psychological factors of participants should be a focus and therefore firms should aim to attract users to join in the online communities.

Many researches have proved that online communities are a promising source of innovations [9, 17, 18]. Furthermore, with the development of the internet, online communities are beginning to be crowded by users' innovations and then how to identify and utilize the user innovations is becoming more and more important for the companies' innovation and new product development process. Therefore, this paper proposed an identifying method on user innovation knowledge in enterprise virtual communities based on weighted knowledge network.

Here, weighted knowledge network is a model of knowledge discovery and representation which is defined as a set of knowledge points, links between knowledge points and weights of them. In WKN, nodes are the knowledge points, links are the relations between these nodes, weights of nodes are the frequency of keywords and weights of links are co-occurrence frequency of keywords. WKN analysis can be expressed as examining the knowledge structure and discovering the relations between different knowledge points [19].

3. RESEARCH METHOD

3.1 Data Mining and Acquisition of Knowledge Points

Data of user innovations which are contained in forum posts will be mined from a selected enterprise virtual community by the information tool LocoySpider. The acquired data sets are expressed as the following formulation:

$$P = \{p_1, p_2, \cdots, p_a\} \tag{1}$$

Based on the data sets, keywords and their frequencies will be figured out and calculated by word segmentation tools. After pre-processing, keywords set and frequency set are acquired respectively. From the knowledge network theory, each keyword represents a knowledge point (K), thus, the keywords mined from the post sets could be regarded as knowledge points in the knowledge networks. Then the knowledge set and the knowledge frequency set are respectively expressed as K and W(K) which are formulated as followed:

$$K = \{k_1, k_2, \cdots, k_n\} \tag{2}$$

$$W(K) = \{w_1, w_2, \cdots, w_n\} \tag{3}$$

Here, knowledge points (K) are selected from high frequency keywords based on the common selection rules of high frequency words which meet the following formulation:

$$\sum [w_i \mid w_i > w_0] \Big/ \sum w_i \approx 0.3 \tag{4}$$

3.2 Acquisition of Knowledge Links and their Weights

In this paper, the edges of knowledge points (R) are their co-occurrence relationships. Thus, the edge set is established according to the co-occurrence of knowledge points in posts. If knowledge points k_i and k_j co-exist in a post, the edge of these two points is expressed as $r_{ij} = 1$, and $r_{ij} = 0$ means there is no co-occurrence between k_i and k_j. According to the frequency of their co-occurrence relationships, the weight of each knowledge edge is calculated as Q (E). Therefore, the sets of knowledge edges and their weights are calculated and expressed as followed:

$$R = \{(r_{ij} = 1) \mid i, j = 1, 2, \cdots, n\} \tag{5}$$

$$W(R) = \{w(r_{ij}) \mid i, j = 1, \cdots, n\} \tag{6}$$

3.3. WKN Model of User Innovations

According to formulations (2-6), the weight of a knowledge point is based on the keywords frequency, which represents the importance of each knowledge point. Since the knowledge points are equivalent to user innovation knowledge points, the weight of knowledge point represents the importance of user-innovations. While the weight of knowledge link is based on the frequency of co-occurrence between knowledge points which represents the closeness between knowledge points and thus represents the closeness between user-innovation knowledge. In other words, the weights of knowledge points and links reflect the trend of user innovation to some extent. Therefore, we can establish the weighted knowledge network model of user innovations as followed:

$$WKN = \{K, W(K), R, W(R)\} \tag{7}$$

4. KNOWLEDGE DISCOVERING ON WKN

4.1. The Method to Identify User-innovation Knowledge in WKN

The WKN model is composed of knowledge nodes, edges and the weights of nodes and edges. The nodes in WKN of user-innovations are keywords which represent the innovation points of a product such as "camera", "function" and "message" in mobile phone, or innovation activities to these points such as "delete" and "fast". The nodes with high weight usually represent important user-innovation points or activities. The higher the weight of the node is, the more important the innovation point or activity is. However, a single node is inadequate to be regarded as the innovation knowledge. Only connecting the innovation points, activities, or points and activities can it be viewed as innovation knowledge, such as "camera function" and "fast delete of messages". Therefore, besides the weight of nodes, the links between nodes should also be considered to identify the hidden innovation knowledge created by users as well.

Since nodes in a network are connected by their edges, there will be several nodes of groups and subgroups with closer relationships such as n-Clique and n-Clan (shown as Fig. **1**).

Here, a clique is defined as a sub-set of actors more closely tied to each other than they are to actors who are not part of the group [20][21]. In social network theory, cliques are regarded as the closest groups of nodes. Based on the definition above, we define a clique in WKN as a close knowledge combination which is composed of more than two knowledge nodes and each two nodes have one link.
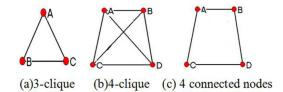

Fig. (1). Nodes combination samples.

Furthermore, edges in WKN have their weights. When edge weights reach or exceed a threshold, the edges can be seemed as strong ties, else as weak ties. If edges in a clique have strong ties, which imply the corresponding innovation activities occur many times, we can say that the clique is of high importance, and it will be identified as a kind of user innovation knowledge.

Therefore, in this paper, we will identify and discover user innovation knowledge on the basis of cliques and tie strengths (*i.e.*, edge weights).

4.2. Identification of Core innovation Knowledge Based on Strong Ties

4.2.1. Sub-Network Construction based on Strong Ties

Setting w_1 as the threshold value of strong ties, which can be set by researchers based on the characteristic of data and research needs. Then the strong ties (R') in WKN will be expressed as followed:

$$R' = \{(r_{ij} = 1) \mid i, j = 1, \cdots, n; w(r_{ij}) \geq w_1\} \tag{8}$$

and the sub-network based on strong ties can be constructed as:

$$WKN' = \{K', W(K'), R', W(R')\} \tag{9}$$

Here, K' is the node set where nodes are connected by edges of R'. $W(K')$ and $W(R')$ are the weights set of nodes and edges in K' and R' respectively. From the formulations, we can see that WKN' is a sub-network of WKN without isolated nodes.

4.2.2. Core Innovation Knowledge Discovering based on Strong Tie Cliques

Based on the WKN', user-innovation knowledge will be identified under the following rules:

Rule 1: Analyze the structure of WKN', and identify all the cliques in the WKN'. Each clique represents a kind of user-innovation knowledge.

A clique here is a maximal complete sub-graph which is composed of at least 3 nodes [20]. The definition shows that there is an edge between each two nodes, thus, cliques are the closest node groups in networks. The strong tie between nodes in cliques indicates the frequent occurrence of the cliques. This indicates that knowledge expressed by cliques with strong ties are the the most important knowledge types in user innovations, in the paper we call them core innovation knowledge.

Every identified clique will be expressed simply by the set of nodes included in each clique. For example, one 3-clique composed of k1, k2, k3 can be expressed as {k1, k2, k3} and one 4-clique composed of k1, k2, k4, k5 will be expressed as {k1, k2, k4, k5}. Then the expression of the set of core innovation knowledge is as below:

$$M_1 = \{(C^1_i) \mid i = 1, 2 \cdots m\} \tag{10}$$

4.3. Identification of Developing Innovation Knowledge Based on the Combination of Strong and Weak Ties

4.3.1. Construction of Sub-Networks Considering Weak Ties

Here, a weak tie means lower value of edge weight which is lower than the threshold value of strong ties but still higher than a certain value. Similar to the construction of sub-network based on strong ties, the sub-network of weak ties (WKN'') is represented as followed:

$$WKN'' = \{K'', W(K''), R'', W(R'')\} \tag{11}$$

Here, K'' is the node set where nodes are linked by links of R''. $W(K'')$ and $W(R'')$ are the weights set of nodes and links in K'' and R'' respectively. Thus, WKN'' is a sub-network of WKN without isolated nodes as well and additionally $WKN'' \supset WKN'$. The edge set R'' can be represented as below:

$$R'' = \{(r_{ij} = 1) \mid i, j = 1, \cdots, n; w(r_{ij}) \geq w_2\} \tag{12}$$
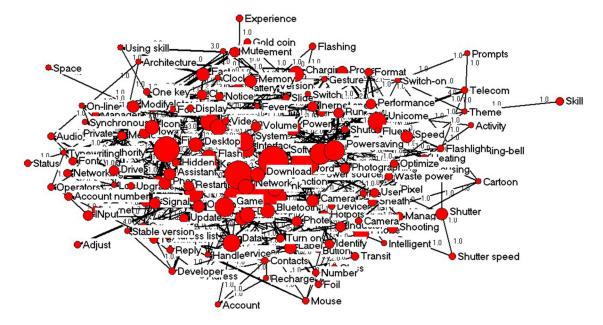
Fig. (2). WKN of user innovation knowledge.

Here w_2 is the threshold value of weak ties which is set by the researcher according to the research needs.

4.3.2. Developing Innovation Knowledge Discovering based on the Cliques with Strong and Weak Ties

In this step, we only focus on the cliques combined of the strong and weak ties, *i.e.*, cliques in strong ties will not be identified.

Rule 2: Analyze the structure of WKN'' and then figure out the cliques combined of strong and weak ties. In other words, the identified cliques should meet the following requirement: at least one edge is strong tie, *i.e.*, it has at least one edge with the weight higher than w_1, others higher than w_2.

The combination of strong and weak ties reflects the growing structure of the innovation knowledge because the weak edges are probably growing into strong ties. Therefore those cliques are defined as the developing innovation knowledge.

The set of developing innovation knowledge is represented as followed:

$$M_2 = \{(C^2_i) \mid i = 1, 2 \cdots l\} \tag{13}$$

4.4. Identification of Potential Innovation Knowledge Based on Weak Ties

4.4.1 Construction of Sub-Networks Only based on Weak Ties

In this step only weak ties are considered. Based on the WKN, we remove all the edges of cliques contained in M_1 and M_2, then delete isolated nodes. The remaining nodes and edges form a sub-network only based on weak ties, which can be represented as below:

$$WKN''' = \{K''', W(K'''), R''', W(R''')\} \tag{14}$$

Here the edge set R''' can be represented as below:

$$R''' = \{(r_{ij} = 1) \mid w_1 > w(r_{ij}) \geq w_2\} \tag{15}$$

4.3.2. Potential Innovation Knowledge Discovering based on the Weak Tie Cliques

In WKN''', there are still many cliques which cannot be viewed as user-innovation knowledge because their edge weights are not higher enough. However, considering the closeness between knowledge nodes in cliques, the ties in this kind of clique is easier to be strengthened. Therefore, we can get the following rule:

Rule 3: Identify the cliques from WKN'''. The identified cliques which are based on weak ties represent potential innovation knowledge for enterprises.

Similarly, the set of potential innovation knowledge is represented as follows:

$$M_3 = \{(C^3_i) \mid i = 1, 2 \cdots l\} \tag{16}$$

5. XIAOMI COMMUNITY CASE

In this part Xiaomi community is taken as the case to verify the feasibility and validity of the proposed methods.

Xiaomi community is the sales, promotion and after service platform for the Xiaomi company and gradually becomes the core of the company's business mode. Furthermore, the community offers a communication platform for users in which users can interact with other users and involve in some innovation activities of Xiaomi Company. Xiaomi community is one of the famous and active user innovation communities with large amount of community members. Up to April 2014, there are more than 226 million posts contributed by users with the amount of 16.83 million community members. The active behavior of users in the community provides a valuable and convenient platform for companies to catch user innovation knowledge.
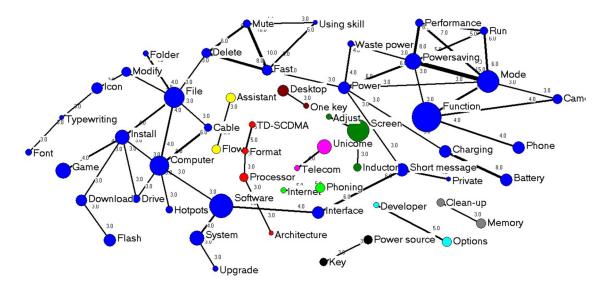
Fig. (3). Sub-network based on strong ties.

5.1. Data Mining and Pre-Processing

In this part, the innovation forum of Xiaomi community was selected out to get user innovation data.

First, LocoySpider 8.2 was used to mine the forum posts developed by users. In this innovation forum, we totally got 3586 posts which was produced and not closed before the date of mining, April 3rd 2014. Next was the pre-procession of these posts. In this step, repeated posts and those with lower browse points of 300 was removed for the convenience of research. After that, we totally got 1154 processed posts for the next processing. Second, keywords of each post were acquired *via* the help of word segmentation tool and totally got 5770 keywords. Third, high frequency words were selected after removing some stop words and calculating the frequency of remaining keywords based on the common rules of high frequency words selection. Then totally we got 150 high frequency words with the total frequency of 1746. The frequency value of all the high frequency words was higher than 5.

5.2. Construction of WKN Model

Based on the high frequency words and their co-occurrence relations, the sets of knowledge nodes, knowledge edges, weights of nodes and edges were acquired so as to construct the WKN model for user innovation knowledge. The network graph was constructed *via* Ucient 6 (see Fig. 2).

Fig. (2) is the WKN graph of user innovations acquired from Xiaomi community which is composed of 150 nodes (the 150 high frequency words) and their co-occurrence relations. The size of each node and each edge represents the weight of the node and the edge respectively.

5.3. Identification of Core innovation Knowledge in Xiaomi Community

5.3.1. Sub-Network Construction

The weight value of 3 was selected as the threshold value of strong ties to construct the sub-network of WKN' (see Fig.

3). The sub-network was composed of 60 nodes and the weight value of edges were equal to or higher than 3.

From the Fig. (3), we can see that this sub-network is divided into several components with different colours of the nodes. Among the components, the biggest one with blue nodes consists of most of the nodes and shows different knowledge patterns which is going to be discussed in next part.

5.3.2. Core Innovation Knowledge Discovering based on Strong Tie Cliques
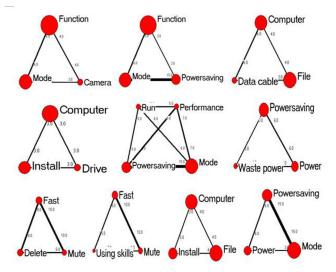

Fig. (4). Core innovation knowledge.

Based on the sub-network of strong ties, cliques were figured out and the set of core innovation knowledge in Xiaomi community was obtained and can also be expressed as sub-network of WKN as followed (see Fig. 4).

From the WKN graph, we can clearly see that the core innovation knowledge in Xiaomi mobile phone focus on power consumption of battery and the running mode of power saving, connection between computer and mobile
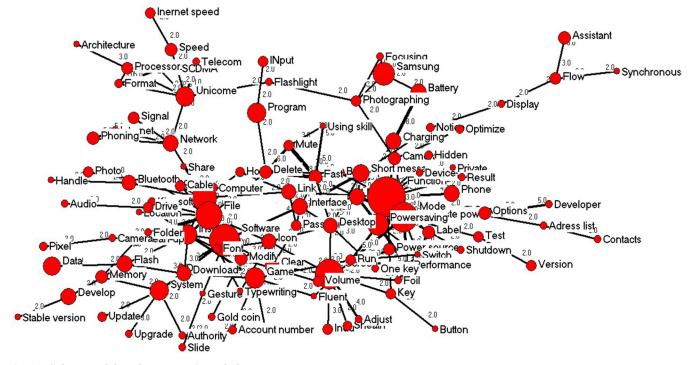
Fig. (5). Sub-network based on strong & weak ties.

phones, fast using skills such as fast mute and delete, and the camera functions. The company should keep eyes on these core innovation knowledge because they embody high values in product innovations.

5.4. Identification of Developing Innovation Knowledge in Xiaomi Community

5.4.1. Construction of Sub-Network on the Combination of Strong and Weak Ties

In this part, we choose 2 as the threshold value of weak ties to construct the sub-network of WKN''. The graph of WKN model was as followed (see Fig. **5**).

5.4.2. Developing Innovation Knowledge in Xiaomi Community

According to rule 3, the cliques were figured out based on the combination of strong and weak ties in the sub-network constructed above. And the set of developing innovation knowledge in Xiaomi community was obtained. Totally, there are 13 developing innovation knowledge identified and constructed as followed WKN graph (see Fig. **6**).

From the WKN graph (Fig. **6**), we can see the growing characteristics of this innovation knowledge. For example, the developing innovation knowledge of "running mode of power-saving" for mobile phone formed by nodes "mode", "power saving", "run", "performance" and "switch" was developed from the core innovation knowledge formed by the 4 including nodes except "switch". This implies that core knowledge modes are likely to become developing knowledge modes and then probably become other core knowledge modes. The company should make good use of these different innovation knowledge in the product innovation activities and new product development.
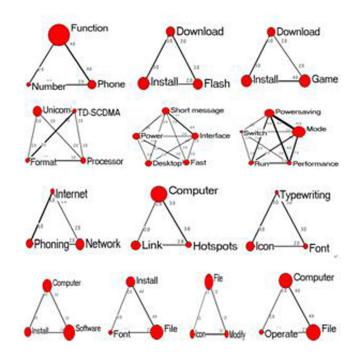

Fig. (6). Developing innovation Knowledge.

5.5. Identification of Potential Innovation Knowledge in Xiaomi Community

5.5.1. Construction of Sub-Networks Only based on Weak Ties

In this part, we still choose 2 as the threshold value of weak ties to construct the sub-network of WKN'''. Since the process and the graph are similiar to those of above, we omit the details in the paper, and the graph is omitted.
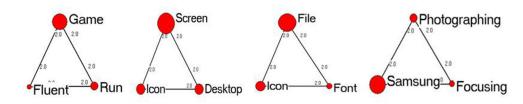
Fig. (7). Potential innovation knowledge.

5.5.2. Potential Innovation Knowledge Discovering in Xiaomi Community

According to rule 4, the cliques in the remaining sub-network were identified (see Fig. **7**).

The WKN graph above shows the 4 potential innovation knowledge. From the graph, we can see that the setting of the photograph focus; the games' running fluency; the screen and desktop icon, and the setting of files icon and script are the four potential innovation knowledge in Xiaomi mobile phone. The company needs to detect the changes of these potential knowledge.

CONCLUSION

This paper proposed a discovering and analyzing method to identify the user-innovation knowledge in enterprise virtual communities. In this method, WKN model and social network analysis were utilized to identify the three kind of innovation knowledge created by user including the core innovation knowledge, the developing innovation knowledge and the potential innovation knowledge. The identifying method is more clearly and accurately to acquire and integrate innovation knowledge contributed by users in virtual communities. The paper suggests a more effective tool for companies to discover and utilize user innovations in virtual communities.

However, there were still some limitations in this study, for example, since the work was based on web-mining, the accuracy of the mining results was hard to be verified. Furthermore, the explanation and discussion of the analysis results need related product knowledge to ensure the accuracy and values. All these limitations may be improved in the future studies.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

[1]   R.J. Füller, M. Bartl, H. Ernst, H. Muhlbacher, "Community based innovation: how to integrate members of virtual communities into new product development", *J. Electronic Commerce Research*, 2006, vol. 6 No. 1, pp. 57-73

[2]   V.H. Eric, and J. Chen, "The major shift towards user-centred innovation", *J. Journal of Knowledge Based Innovation in China,* 2009, 1(1), 16-27

[3]   G.L. Urban, and V.H. Eric, "Lead user analyses for the development of new industrial products", *J. Management Science,* 1988, 34(5), 569.

[4]   S. Shah, "Sources and Patterns of Innovation in a Consumer Products Field: Innovations in Sporting Equipment", *MIT Sloan School of Management Working Paper* No. 4105, March, 2000.

[5]   J.L. Enos, "Petroleum Progress and Profits: A History of Process Innovation", *MIT Press*, Cambridge MA. 1962

[6]   C. Freeman, "Chemical process plant: innovation and the world market", *J. National Institute Economic Review*, 1968, Vol. 45, pp. 29-57.

[7]   K. Pavitt, "Sectoral patterns of technical change: towards taxonomy and a theory", *J. Research Policy,* 1984, Vol. 13 No. 6, pp. 343-73.

[8]   V. H. Eric "Lead users: A source of novel product concepts", *J. Management Science,* 1986, 32(7), 791.

[9]   P.D. Morrison, J.H. Roberts, V.H. Eric, "Determinants of user innovation and innovation sharing in a local market", *J. Management Science,* 2000, Vol. 46 No. 12, pp. 1513-27.

[10]   C. Lettl, C. Herstatt, H.G. Gemuenden, "Users' contributions to radical innovation:evidence from four cases in the field of medical equipment technology", *J. R&D Management.*, 2006, vol. 36 No. 3, pp. 251-72.

[11]   C. Herstatt, V.H. Eric, "From experience: Developing new product concepts *via* the lead user method: A case study in a "low tech" field", *J. Product Innovation Management.* 1992, 9(3)213-221.

[12]   E.L. Olson, G. Bakke "Implementing the lead user method in a high technology firm:A longitudinal study of intentions versus actions", *J. Product Innovation Management.* 2001, 18(2)388-395.

[13]   N. Franke, S. Shah, "How communities support innovative activities: An exploration of assistance and sharing among innovative users of sporting equipment", *Sloan Working Paper* #4164. 2001

[14]   R.J. Füller, M. Bartl, H. Ernst, H. Muhlbacher, "Community based innovation: a method to utilize the innovative potential of online communities", *Proceedings of the 37th Hawaii International Conference on System Sciences*, IEEE, Big Island, HI. 2004

[15]   R. Kozinets, "The field behind the screen: Using netnography for marketing Research in Online Communications", *Journal of Marketing Research,* 2002, 39(1), 61–72.

[16]   K.M. Chu, H.C. Chan, "Community based innovation: Its antecedents and its impact on innovation success", *J. Internet Research,* 2009, 19(5), 496-516.

[17]   E. Verona, N. Sachs-Ericsson, T.E. Joiner, "Suicide attempts associated with externalizing psychopathology in an epidemiological sample", *J. American Journal of Psychiatry,* 2004, Vol. 161, pp. 444-51.

[18]   V.H. Eric, R. Katz, „Shifting innovation to users *via* toolkits", *J. Management Science,* 2002, 48(7), 821-833.

[19]   Y.J. Xi, Y.Z.Dang, "The Discovery and Representation Methods of Expert Domain Knowledge Based on Knowledge Network", *J. Systems Engineering*, 2005, 08:110-115.

[20]   J. Scott, "Social Network Analysis: A Handbook", *Sage*, London. 2000.

[21]   L. Boyer, R. Belzeaux, K. Baumstarck-Barrau, J. Samuelian, O. Maurel, "A social network analysis of healthcare professional rela-tionships in a french hospital", *J. International Journal of Health Care Quality Assurance,* 2010, 23(5), 460-9.