Winter 12-2-2007

Knowing who to know in Knowledge Sharing Communities: A Social Network Analysis Approach

Jyun-Cheng Wang

Ming-Jiin Chiang

Jung-Chang Ho

Wen-Tai Hsieh

I-Kun Huang

Proceedings of The Seventh International Conference on Electronic Business, Taipei, Taiwan, December 2-6, 2007, pp. 345-351.

KNOWING WHO TO KNOW IN KNOWLEDGE SHARING COMMUNITIES: A SOCIAL NETWORK ANALYSIS APPROACH

Jyun-Cheng Wang, National Tsing Hua University, Taiwan, jcwang@mx.nthu.edu.tw
Ming-Jiin Chiang, National Chung Cheng University, Taiwan, mingjiin@mis.ccu.edu.tw
Jung-Chang Ho, Taiwan Semiconductor Manufacturing Company Ltd, Taiwan,
Wen-Tai Hsieh, Institute for Information, Taiwan, wentai@iii.org.tw
I-Kun Huang, Institute for Information, Taiwan, yikuan@iii.org.tw

ABSTRACT

Information stored in online communities consist not only knowledge contents, but also the information of knowledge providers and searchers' connective relationships, and network structures. Online Communities provide effective platforms for interaction and play pivotal roles in making provision for the basis of analysis as all the ask-response paired relationships are automatically recorded. This paper demonstrates how to apply social network analysis to analyze the interaction data for generating the "role information" of the knowledge searchers and providers. Integrating concepts of uncertainty in knowledge searching and sociometric used in social network analysis, we develop a mechanism for role matching in knowledge search for each questions posed. Roles identified in this approach including central, network entrepreneur (e.g. spanning structural holes), neighboring mediate (e.g. knowledge gate keeper), and resource competitor (e.g. structural equivalent players). The result is demonstrated and visualized in a web-based community platform and tested in a real-world programmer forum-based community.

Keywords: Social network analysis, role analysis, knowledge network, knowledge community.

INTRODUCTION

Studies of knowledge management show that the success of knowledge transfer lies in neither communication systems nor documents, but in social relationships [1][3][6][7][16]. Different social relations tend to result in different consultations. All these point to the fact that knowledge can be better captured by shifting focus from the face value of the content itself to the social network where knowledge is embedded.

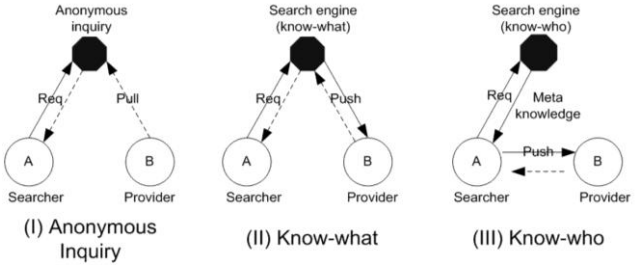

Figure 1. Knowledge Inquiry Models with computer-mediated Interaction

In a knowledge intensive online virtual community, such as a technical discussion forum, knowledge sharing usually starts with knowledge inquiry took place in an online platform, which can take three models as depicted in Figure 1. The first one illustrates anonymous inquiry where inquirers, or searchers, post questions in a forum waiting for answers from volunteers. This mode simply works in a post-and-wait manner. The second one, known as "Know-what" mode, describes most situations where huge volumes of documents or knowledge repositories are stored in a database ready to be retrieved by keyword matching or other more sophisticated yet similar techniques. An inquirer can only obtain what is stored in the database, and one may have difficulty in evaluating the quality of the information acquired.

The third approach is to make provision of linking knowledge contents to those experts who are also community members. This is the "know-who" mode, illustrated in the right most panel of the Figure 1. In this mode, the connections between searchers and knowledge providers can be established. The main task in this model is to have the knowledge of who knows about the answer, and make it available to the searchers. Meanwhile, searchers in this model will have to provide something more than keywords of their questions. They also need to reveal their level of knowledge, so the system can learn to recommend someone suitable for conducting dialogues. As a result, communities will become a platform for knowledge sharing and bring not only knowledge but also the information of provider and searcher's relative network positions, connective relationships, and network structure. For example, knowledge search engines in this model can show that people in what position are inclined to help, or whose knowledge source are more close to searchers' background.

Communities based on the "know-who" model will be more effective than the "know-what" model by facilitating knowledge transfer and enhancing experiences and value exchange, and thus help accomplish knowledge sharing through social interaction. At the end, knowledge is in the head of people [19]. From the point of avoiding "free-riders", this type of interaction via social embeddedness will help reduce improper responses [2][14] and boost sense of community and is favorable for the formation of social norm [9].

The purpose of this study is to realize the concept of implementing a "know-who" knowledge sharing platform by analyzing role information and evaluating relationships among community members from community interaction data. By utilizing the role information, we build a community platform for automatically identify the proper individuals for a query. In finding the proper person to answer a question, we consider issues from such different aspects as knowledge content, social context, and personal knowledge.

LITERATURE REVIEW

Knowledge Sharing

Knowledge generation includes not only objective processes of transferring information into knowledge through comparison, cause and effect analysis, interlink and communication but also subjective process of generating personal interpretation, through experience, reality, judgment, law, institution, etc. [8]. As knowledge is highly personalized, it has to be expressed through personal experience, impression, practiced skill, culture, or shabit [15][18]. Knowledge sharing usually starts with some kind of inquiry and search. However, answers can be provided in three different forms, as shown in Table 1. In this classification schemes, these search engine may either be focused on knowledge contents, social context, and personal preferences and profiles. Several packages or tools available for the different types of searching and browsing are listed in Table 2.

Table 1. Classifications of Search Engine

Search Mode	Cues for Search	Relevant Techniques
Knowledge content focus	Domain-specific Knowledge, Ontology, Knowledge-base, Thesaurus	Natural language, Semantic web, key-graph
Social-context focus	Feedback, Conformity, Citation relationship, Recommendation, Social network	PageRank, Citation Analysis, Social network analysis
Personal preferences focus	Personal searching history, Preference	Agent, personal KM system

Table 2. Tools for Knowledge Browsing

Research	Objective	Contribution
Sack (2000) [21]	Conversation map for Newsgroup	Visualize very-large-scale conversation map of USENET by integrate social network and semantic network
Merali et al. (2001) [12]	Knowledge capture and utilization in Virtual Communities	Jasper II (a knowledge sharing environment) use information agent for sharing knowledge from a number of internal and external sources.
Smith et al. (2001) [23]	Persistent Conversation on newsgroup	A set of tools for visualization of the structure of discussion threads and the pattern of participation within the discussions
Lin et al. (2003) [11]	Knowledge map creation and maintenance in Virtual Communities	Generate Knowledge map by text-mining on documents collected from teachers' cyber community

When viewing knowledge seekers and providers as knowledge buyers and sellers, knowledge market is like the usual market in the sense that certain uncertainties involved. Just as a purchase decision will consider reputation of stores besides goods and prices to control the risk of purchase, knowledge search could reduce uncertainty in knowledge seeking by taking account of not only knowledge content but also knowledge owner and relationship between each other [8]. According to Podolny [17], there exist two types of uncertainty in seeking knowledge in a knowledge market; one is high ego-centric uncertainty and the other high alter-centric uncertainty. The former uncertainty is the risk in searching for transaction chances; the latter is the doubt about the quality of the acquired knowledge.

To avoid the uncertainty existing in knowledge markets, social network can play the role of pipe and prism [17]. Pipe effect refers to the fact that social networks can extend human relations and connects ways to acquire resources. As the tacit feature of knowledge paralyzes the fluidity [15], one can rely on social networks for identifying the chances to acquire knowledge. The phenomenon of birds of a feather helps reveal that social networks have the capacity to agglomerate the same resources and save search cost. It is evident that social relationship is a significant element of knowledge transfer [5][8][9][20]. The Prism effect refers to the social network functions to filter lower quality while too many choices. This is made possible by exploiting the extension of the trust relationship or observation of other social network structures.

Social networks make provision for successful knowledge transfer between strong ties and weak ties. Communities tend to agglomerate people who have the same interest and construct strong ties for trust-based knowledge exchange. From outside communities, weak ties would bring resources and avoids partition of social network. Knowledge management will rely on both types of ties for transferring useful knowledge. Levin (2002) suggests the combination of weak and strong ties could help highly tacit knowledge transfer through two kinds of trust, affection and competence trust [10]. The existence of trust would be a critical basis for establishing non-money based markets for knowledge transfer. The accumulation of social capital in turn depends on the smooth operations these types of markets.

Role Analysis

To make a community platform suitable for knowledge transfer, role analysis is an important function. Role analysis in socio-metric is used to analyze actors who have similar structures or patterns [24]. The possible attributes for role analysis is listed in Table 3. These methods are derived from mathematic and multiple study fields. Similar positions in SNA reflect the relative status of individuals who are embedded in network having similar relations. Roles could reveal the same pattern

between members or positions. SNA adopts pattern match for role discrimination, which needs network index to find the same pattern role. Therefore, community platform has to provide related index for computing and comparing.

Table 3. Role Analysis

Role Attribute Type	Description	Example
Peripheral attributes	The peripheral attributes of a role are those expectations which members of the society have about a particular role that are insignificant in meeting the obligations implied by the role.	Sex/gender of a physician is peripheral to his/her functioning as a physician.
sufficiently relevant attributes	The sufficiently relevant attributes of a role are those expectations which members of the society have about a particular role that if they are missing, sanctions will be invoked.	If my physician refuses to provide a medical examination when requested to do so, sanctions may be imposed to require the examination to take place.
Pivotal attributes	Pivotal attributes are those which if they are absent, the role is said not to exist.	For instance, someone who purports to be a physician but who has not passed the appropriate medical examinations is not a physician but rather a charlatan.

Keyword Expansion

Two different relations between any pair of tags can be defined by aggregating tagging data. First, two tags are in a "Co-Resource" relation if they are adopted for the same resource. This relation is stronger between tags with more shared resources. Second, two tags used by the same user are in a "Co-User" relation. While the "Co-Resource" relation is most appropriate for establishing a public concept hierarchy, the "Co-User" Relation is most suitable for establishing a private concept hierarchy

As a result, the concept hierarchy can be used to expand keyword. The link of the target keyword and other keywords in the concept hierarchy indicates the relation of two keywords.

SYSTEM ANALYSIS AND DESIGN

The "Role" information

The search for a proper target for knowledge exchange may be subject to a variety of uncertainty. As we mentioned above in the context of Knowledge Exchange Market, this uncertainty includes High Ego-centric Uncertainty and High Alter-centric Uncertainty [17]. The former refers to the uncertainty in discovering the opportunity for exchange, and the latter, the uncertainty about the quality of the knowledge obtained via the exchange.

The "Role Information" is designed to be used in conjunction with other search cues in reducing the uncertainty of knowledge search and increase the success rate of knowledge exchange. One person's expertise level varies as the topic in question changes, or when questioners differ. The "role information" of an individual has to be dynamic and cannot be treated like a static label. It should be determined upon who queries and what is queried. In light of this, we identify ten functional requirements shown in Table 4.

Table 4. Functional Requirements and possible solutions

No	Description of Functional Requirements	Possible Solutions
1	Provision of a list of topics available for search	Use social network analysis for Keywords, and identify the main topics and their keywords
2	Provision of Keywords analysis of the usable topics	
3	Capable of making references to one's closely interacting sub-groups	Revealing the user's position in the social network
4	The number of search results can be adjusted as needed	Targets will be sorted in terms of their relevance to the query
5	Provision of information about the familiarity between the user and members	Use "neighborness" in network structure to represent their relationships
6	Provision of information about the depth of members' expertise	Use "significance" in network structure position to represent the expert's status
7	Provision of information about the breadth of member's expertise	Use "structure hole" in network position to show its variety
8	Evaluation of equivalence in expertise level	Use "structural equivalence" to find status equivalence between users and experts
9	Evaluation of bottleneck in knowledge exchange	Use intermediary analysis techniques to identify knowledge intermediary
10	Allowing users modifying their own experiences information	Availability of users modifying role information

Performing Role Analysis

We conclude from the functional requirements that there are four types of role analyses will contribute to reducing the uncertainty of searchers. These types of role analyses and the corresponding uncertainty they can help reduce is shown in Figure 2. The meaning of each type of analysis is described as follows.

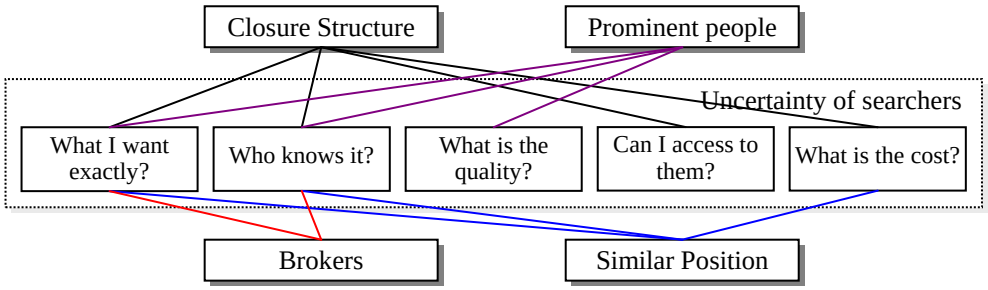

Figure 2. Use of four types of Role Analyses to reduce the corresponding uncertainly

1. Closure Structure: Actors with shared preferences for information sources tend to cluster together and form strong ties, leading to a closure structure. This analysis can be used to measure the closeness in terms of relationships and preferences similarity.
2. Similar Position: This is used to search targets that link to similar actors and are of similar network structures. These targets enjoy common features and are highly substitutable. They tend to be peers in their capabilities and may compete for similar resources.
3. Brokers: This is defined with respect to the position of searchers. Brokers play a critical role in controlling or facilitating the delivery of information. Possible roles may include gatekeepers, inter-mediators, and representatives.
4. Prominent people: Two types of actors are in prominent positions of network structures. One type of the prominent roles carries high centrality and is of high level of influences, e.g. opinion leaders. The other type spans multiple network structure holes, and usually can access various sources of information. This is sometimes called network entrepreneurs.

The four types of analyses are produced by utilizing some of the procedures for calculating indexes that are available in network analysis procedures, such as Distance, Structure Equivalence, Centrality, and Broker.

The Role Analysis Process

Our proposed role analysis is performed on top of the usual keyword-based search. That is, we take as input the resultant lists from keyword-based search and feed them into the role analysis process. There are five steps in this process. We first conduct an Absolute Role Analysis (step 1~3), where roles of prominent status and of spanning structure holes are identified. The next is to do Relative Role Analysis (step 4~5), in which the role information obtained from step 3 will be interpreted from the perspective of the searcher. This process is depicted in Figure 3.

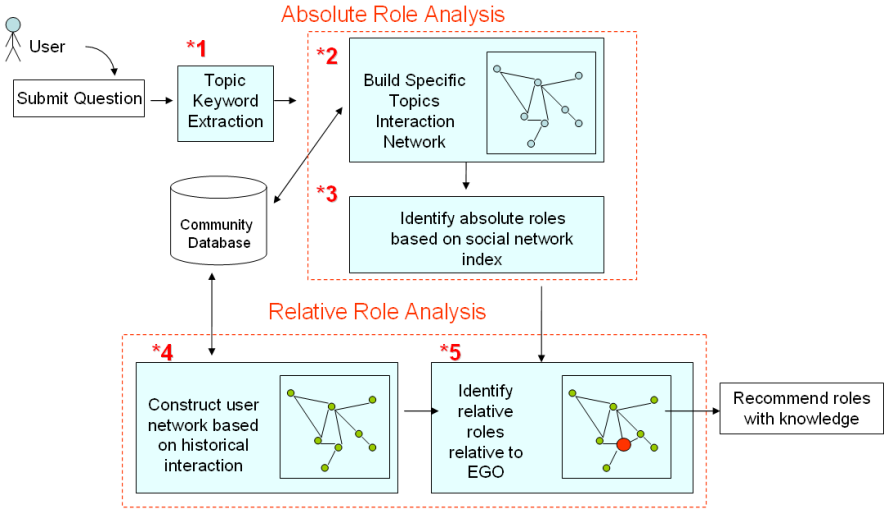

Figure 3. The Role Analysis Process

Based on keyword database and its associated keyword concept space, the system extracts keyword sets from user's question in the first step. On receiving user's queries, system automatically extracts several terms or nouns as the first level keywords. Next, system extends second level keyword associated with first level keywords by referring to the keyword concept space. For example, 'JAVA' is the first level keyword when user submits a JAVA-related question, and 'J2EE' and 'J2SE' are the extended keywords (or second level keywords). The extending action is carried out iteratively by system until the default threshold level reached. The keyword sets, including first level keywords and other level keywords extended by system, were the basis to construct Topics Interaction Network in the second step.

In second step, system looks for actors who had participated in the discussion with keyword sets derived in first step in order to construct Topics Interaction Network. If two users interact in the same topic/keyword, we build edge between them. Thus, system constructs Topics Interaction Network by searching actors who had participated in the discussion of every keyword from the community database.

The system also looks for brokers and prominent people in the Topics Interaction Network in the third step. We draw on the Structural Hole algorithm proposed by Burt to calculate the effectiveness and efficiency for each actor in the Topics Interaction Network, and list N actors (assuming N is the system default number of candidates) with the highest efficiency. An actor with high efficiency is more likely in the position of acting as broker for other actors. Further, system looks for prominent people by calculating degree centrality of each actor in the Topics Interaction Network. An actor with high degree centrality maintains contacts with various other actors and is a people of influence in the network. Thus, system will select n actors (n is the system default candidate) with the highest centrality as the candidate of prominent people.

The main task of the fourth step is extending the Topics Interaction Network into a User Interaction Network. Topics Interaction Network includes experts who are familiar with keyword sets and yet excludes questioner. We want to add actors who ever interact with questioner from community database to form a Topics Interaction Network. The edge in the User Interaction Network indicates the degree of two person's interaction. This network demonstrates those actors who ever interact with questioner and domain experts who ever joined the topic with keyword sets.

Finally, the system looks for the shortest distance between domain experts and questioner. In this final step, the system would search targets that are of similar network structures. We calculate the distance of any two adjacent nodes in the User Interaction Network and identify the shortest path between actors and questioner. The nearer the distance from questioner (or closure structure) indicates that the questioner is either more likely to find experts or the questioner and experts share the same sub-group. The system will select N actors (N being the system default number of candidates) with the lowest distance as the candidate actors of domain experts. Finally, the system will draw on structure equivalence algorithm to calculate the structure equivalence (that is, similar position) between the questioner and other actors. We utilize Euclidean Distance and CONCOR (convergence of iterated correlation analysis) to list the top N actors (N being the system default number of candidates) with the closest structure equivalence. These candidate actors enjoy common features and are highly substitutable. They tend to be peers in their capabilities and may compete for similar resources.

SYSTEM DEVELOPMENT AND AN EXAMPLE

The platform

We develop the system on top of a virtual community support platform. Conceptually, this system follows the IPO structure. At the bottom is the Community Data, which is fed into the middle layer for Role Analysis. Data are parsed and tagged, full-text contents analyzed for terms and keywords, which in turn is used to build a keywords network. Keywords will form clusters for those more likely to appear simultaneously, using the RNM algorithm [13]. This algorithm uses Peer Influence Model for calculation and classification, it is considered both easier and faster than general Partition techniques in Graph theory.

System Testing

One aspect that makes this system distinct from general virtual community platform is the capability of visualization of the recommended experts. Social Network Analysis originates from Graph Theory and shares the basic components of nodes and links. However, in order to make the nuts and bolts of network graph meaningful in presentation, we design the screen to consist of several units to reveal meanings of network pictures. These units include: Information Tagging, Scene Generator, Analysis unit, Operation unit, and Viewpoint unit.

We test our system by analyzing a forum-based community mainly for interactions between programmers. Topics are classified and threaded, and people appear on this community includes some of the well-known figures. The total registered member is more than five thousand people. The forum consists of 48 discussion boards, with topics more than twenty thousand.

After entering a keyword, a searcher can expect the system to return a network graph representing all the people, or experts, who have engaged in discussions of related topics. These "knowers" possess one particular network position that could be close or further away from the searcher. Searchers can traverse on the network graph to find one with desired status. All can be accomplished under the assistance of the system.

The search results shown in Figure 4 show the information presented in the system. The red rectangle at the left hand side represents the position of the searcher. All other actors are drawn with respect to the searcher's knowledge status (novice or experts), so the relative role information can be examined. Clicking on any node in the Tree Layout, the searcher can learn more of the selected actor's information. This information, shown in the right hand side, describes one actor in terms of five types of search uncertainty. The current picture adopts a Tree Layout, the system also provides a Spring Embedding Layout, and is capable of enlarging views.

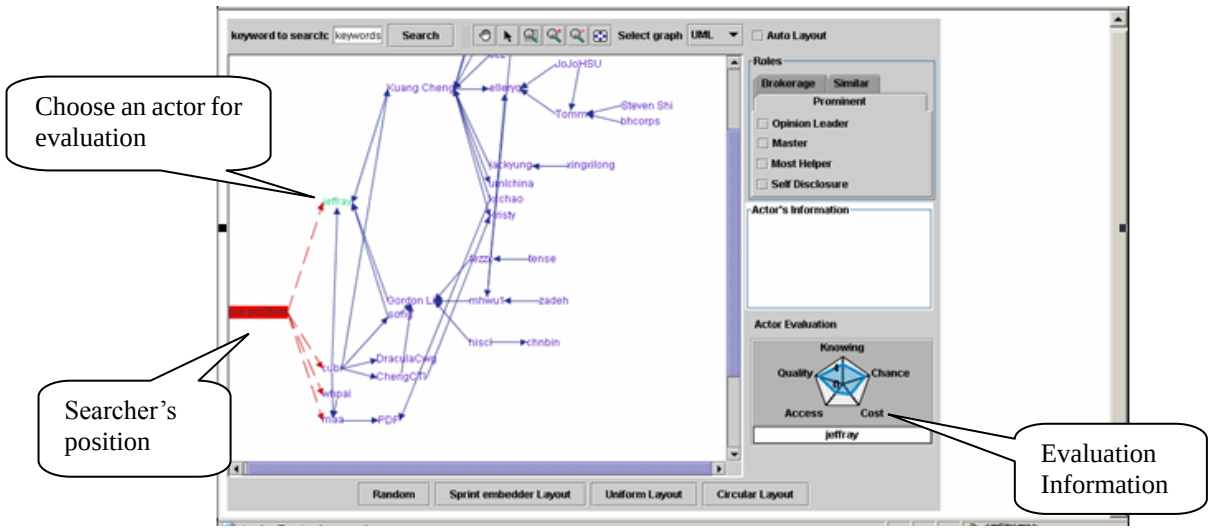

Figure 4. Search Results: Representing in Tree Layout

DISCUSSIONS AND CONCLUSIONS

Discussions

The purposes of this research include are twofold:

1. Model establishment: The system has been successfully finding out the role type which influences knowledge exchange procedure, identifying method for analysis, evaluating the value of a role, and establishing effective role search model.
2. System development: Deploying search techniques based on a community platform, developing the search system which employs the role search model to make possible the personalized knowledge filtering mechanism.

When new media emerges, it usually brings the transformation of related application. Essentially, knowledge search should not only concentrate on knowledge content itself but also has to include social interaction. Prusak has claimed that the value of knowledge depends on the exchangers' social relationships. The emergence of virtual community brings the social interaction to a new level and can be operationalized.

There are some limitations in this study. Member's background in virtual community is far more complex than that in real world. The fluidity and anonymity of membership makes it difficult to distinguish participants' identity [25]. The same individual with multiple identities also makes analysis troublesome.

The other problem is on the possible invasion of privacy or moral issues when it comes to data collection [4]. When apply this system in a real world environment, issues need to be resolved on whose permission do one need before proceeding to use community information? community constructer? administrator? or the person who posts? It is largely an open question and depends on the nature of the information and the situation of use.

Smith (1997) has also pointed to the limitations of online anthropology [22]. There are four types: the lack of generality, the lack of correlation, the lack of historical information, and the lack of scope. Among all these, this research does not escape easily.

Suggestions and Future Study

The development of this type of search system which employs social network analysis techniques is still emerging. This system provides modularized system models and friendly user interface; however, it has yet to be tested in a larger scale site. Further testing of the system can be conducted in a pseudo knowledge market, where knowledge needs to be priced, evaluated, and traded. One strong point of this system is that it can be used for different purposes of knowledge search. The fitness of knowledge search is investigated quite often, and it reveals that content is only one aspect of knowledge. Information techniques can be applied to help expose the various aspects of value.

The focus on role information leading to challenges against the popular approaches of measuring member performance based on the frequencies of posting. Many experts in a field only respond to the most critical questions, while leaving the general problem answered by lower status experts. The recognition of role information can be beneficial to the development of community if used effectively and reduce the adversary bottleneck effects.

Communities can not be easily classified as a hierarchy structure or a market. Rather, it is like an epitome of society. Our experiences in employing Social Network knowledge for the study of knowledge sharing in virtual community stand on a convenient yet solid theoretical ground. We think that it will be a fruitful research direction.

ACKNOWLEDGEMENT

This research was supported by the New Digital Learning Environment Technology Five Years Research Project of Institute for Information Industry and sponsored by MOEA , ROC.

REFERENCES

[1] Argote, L., McEvily, B. and Reagans, R. (2003) "Managing knowledge in organizations: an integrative framework and review of emerging themes", *Management Science*, Vol. 49, No. 4, pp. 571-582.

[2] Beinhauer, M. (2000) "Collective knowledge management via virtual communities", *Proceedings of the 2nd International Conference MITIP 2000*.

[3] Borgatti, S.P., and Cross, R. (2003) "A relational view of information seeking and learning in social networks", *Management Science*, Vol. 49, No. 4, pp. 432-445.

[4] Bruckman, A., Erickson, T., Fisher, D., and Lueg, C. (2000) "Dealing with community data: a report on the cscw 2000 workshop", CSCW 2000 Workshop.

[5] Burt, R.S. (2002) "The social capital of structural holes", *The New. Economic Sociology*, M.F. Guillén, R. Collins, P. England and M. Meyer (eds.), Russell Sage Foundation, New York, pp. 148-192.

[6] Cross, R., Borgatti, S.P., and Parker, A. (2001) "Beyond answers: dimensions of the advice network", *Social Networks*, Vol. 23, No. 3, pp. 215-235.

[7] Cross, R., Rice, R.E., and Parker, A. (2001) "Information seeking in social context: structural influences and receipt of information benefits", *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, Vol.31, No. 4, pp. 438-448.

[8] Davenport, T.H., and Prusak, L. (1998) *Working knowledge: how organizations manage what they know* Harvard Business School Press, Boston, Mass.

[9] Granovetter, M. (1973) "The strength of weak ties", *American Journal of Sociology*, Vol.78, pp. 1360-1380.

[10] Levin, D.Z., Cross, R., and Abrams, L.C. (2004) "The strength of weak ties you can trust: the mediating role of trust in effective knowledge transfer", *Management Science*, Vol. 50, No. 11, pp. 1477-1490.

[11] Lin, F.-R., and Hsueh, C.-M. (2003) "Knowledge map creation and maintenance for virtual communities of practice", P*roceedings of the 36th, Hawaii International Conference on System Sciences*, Big Island, Hawaii, pp. 1-10.

[12] Merali, Y. and Davis, J. (2001) "Knowledge capture and utilization in virtual communities", *Proceedings of the international conference on Knowledge capture*, Victoria, British Columbia, Canada, pp. 92-99.

[13] Moody, J. (2001) "Peer influence groups: identifying dense clusters in large networks", *Social Networks*, Vol.23, No. 4, pp. 261-283.

[14] Neus, A. (2001) "Managing information quality in virtual communities of practice", *Proceedings of the 6th International Conference on Information Quality at MIT*, Sloan School of Management, Boston, MA.

[15] Nonaka, I. and Takeuchi, H. (1995) *The knowledge-creating company: how Japanese companies create the dynamics of innovation* Oxford University Press, New York.

[16] Papakyriazis, N.V., and Boudourides, M.A. (2001) "Electronic weak ties in network organizations", *4th German Online Research Conference*, Goettingen, Germany.

[17] Podolny, J.M. (2001) "Networks as the pipes and prisms of the market", *American Journal of Sociology*, Vol. 107, No. 1, pp. 33-66.

[18] Polanyi, M. (1966) *The tacit dimension anchor day books*, New York.

[19] Prusak, L. (2002) "Second generation knowledge management", *Capability Magazine*.

[20] Rogers, E. (1995) *Diffusion of innovations* Free Press, New York.

[21] Sack, W. (2000) "Conversation map: an interface for very-large-scale conversations", *Journal of Management Information Systems*, Vol. 17, No. 3, pp. 73-92.

[22] Smith, M.A. (1997) "Measuring and mapping the social structure of Usenet", *17th Annual International Sunbelt Social Network Conference*, San Diego, USA.

[23] Smith, M.A., and Fiore, A.T. (2001) "Visualization components for persistent conversations", *Proceedings of the CHI 2001 Conference on Human Factors in Computer Systems*, Seattle, Washington.

[24] Wasserman, S., and Faust, K. (1994) *Social network analysis: methods and applications*, Cambridge University Press.

[25] Wellman, B., Salaff, J., Dimitrova, D., and Garton, L. (1996) "Computer networks as social networks: collaborative work, telework, and virtual community", *Annual Review of Sociology*, Vol. 22, No. 4, pp. 213-238.