Analyzing Community Knowledge Sharing Behavior

Styliani Kleanthous
School of Computing, University of Leeds
stellak@comp.leeds.ac.uk

Vania Dimitrova
School of Computing, University of Leeds
vania@comp.leeds.ac.uk

ABSTRACT

The effectiveness of support provided to virtual communities depends strongly on *what we know* about a particular community of people and *in what area*s the community needs support. For this purpose, we have developed algorithms to automatically detect patterns that point out support needed for the effective knowledge sharing in a closely-knit virtual community. The automatic detection of problematic areas enables the development of intelligent solutions aimed at different community members, but, at the same time, supporting the knowledge sharing practice of the community *as a whole*. This paper will provide a brief overview of our approach which applies user modeling techniques, and will focus on graph based algorithms developed for detecting patterns that can be used to provide intelligent support to virtual communities.

General Terms

Algorithms, Human Factors

Keywords

Community knowledge sharing, closely-knit communities, adaptive support, community modelling

1. INTRODUCTION

Virtual communities (VC), allow people to gather together and share knowledge disregarding time and space differences. In a broad sense, VC vary from fairly large, loosely structured to small, closely-knit ones. Studies have shown that having technology and people present does not guarantee efficient knowledge sharing and sustainability in a virtual community [5]. Appropriate support is needed to facilitate the functioning of a community where the members actively engage and share knowledge effectively [6]. Computer supported cooperative work (CSCW) research has exploited different approaches to support people working in shared workspaces, such as visualizations, notifications and awareness techniques. In most of the cases, these approaches are developed ad-hoc without following particular organization or psychology theory, have a problem-oriented focus, or are based on ethnographic methods [3]. Although these approaches significantly advanced knowledge in supporting communities, they are usually difficult to generalize due to their lack of theoretical underpinning [3].

This research aims at supporting effective knowledge sharing and sustainability in VC by providing community-tailored support based on theoretical foundations and analysis of log data, collected from a real virtual community. Support will be targeting the individual but aiming to support knowledge sharing in the community as a whole. We consider closely-knit virtual communities for knowledge sharing which are composed by participants with common interests, commitment to the sharing of information, resources and generation of new knowledge, and equal membership inside the community. Sustainability in VC requires members to actively engage in both contributing to and benefiting from the community's knowledge for as long as the community is active.

Closely-knit VC usually exist in relatively well-defined organizational or educational settings, and can share common characteristics with teams. Following research in organizational psychology [7], we have identified several processes important for effective team functioning which can be applied to VC and can be examined or facilitated by analyzing community log data. These processes include:

Transactive Memory (TM): members are aware how their knowledge relates to the knowledge of the others [17].

Shared Mental Models (SMM): members develop a shared understanding of the key processes and the relationships that occur between them in the community.[13]

A good transactive memory system and the establishment of shared mental models are both important for healthy collaboration between people.[8]

Cognitive Centrality (CCen): members who hold strong relevant expertise can be influential; it has been shown that members of effective communities gradually move from being peripheral to becoming more central and engaged in the community [12].

Algorithms have been defined and developed in order to extract a community model (CM) based on the above process. The CM in brief consists of: (a) individual user models, (b) a model of the semantic relationships between community members, (c) a list of cognitively central members, (d) a list of popular and peripheral topics in the community, and (e) the community context defined by a subject-specific ontology [10].

The community model is presented in [11] and applied to tracking data from a real community to get an understanding of what is happening in the community, and to identify what support can be provided to improve the functioning of the community. Furthermore, [11] uncovered knowledge sharing behavior patterns that needed further investigation. These patterns were manually detected by examining the community model with appropriate visualization tools. To enable the automatic detection of such patterns, we have developed graph-based algorithms that analyze the community model and identify problematic areas in the community where adaptive support would be required. This paper presents the algorithms for detecting community knowledge sharing patterns, applies these algorithms to a real community, and discusses how the patterns detected can be used to improve the TM system, to help establish SMM and to enhance collaboration in the community.

With the research presented in this paper we want to answer the following questions:

Q1: What patterns can be detected, with respect to TM, SMM and collaboration within VC?

Q2: Can we use the detected patterns to improve the knowledge sharing process and the sustainability of VC?

Q3: How can the knowledge sharing behavior of members affect the development of a good TM system, the establishment of SMM, and the development of collaboration?

The next section will position our research in the relevant body of research by compare with related CSCW approaches. Section 3 will discuss what patterns should be discovered, while Section 4 will introduce the main definitions used in this paper. Section 5 will give a detailed description of the graph algorithms developed for detecting community knowledge sharing patterns, and will explain their importance with regard to TM, SMM, and CCen). Section 6 will present the application of these algorithms in a study with tracking data from an existing community and the results will be discussed in section 7. Finally, Section 8 will conclude and point at future work.

2. RELATED WORK

There is a growing interest in providing adaptive support for teams, groups and communities. A well-researched area is that of expertise finding. Different tools and algorithms have been developed to support people in locating expertise on a specific subject inside small or large groups or VCs [1, 15, 18]. Our approach relates to identifying expertise. However, in addition to identifying the interests and expertise of community members, we derive a person's influence in the VC based on the relationships he/she has developed with others. The ultimate goal is to benefit the VC as a whole by encouraging cognitively central members to engage in interactions with peripheral members and help them better integrate in the community.

Visualisation techniques are another approach for providing awareness of what is happening in a community, and thus, supporting participation and collaboration in a VC. For example, graphical representations are used to make people aware of the relevance to the activity or to the position of a particular member in the group [9] or to show the status (or popularity) of a resource [2]. The key limitation of visualisation techniques is their passive influence on the functioning of the community, e.g. while examining graphical representations members may not be able to see how their contribution could be beneficial for the community. In contrast, our approach proposes the use of an extended community model to *automatically detect* problematic cases which can be used to decide when and how to intervene and offer support to improve the knowledge sharing and sustainability of the whole community.

Notification of actions happening in the VC space has also been researched in CSCW. Systems have been developed to support social awareness (who is present) and task awareness (what activities have happened) [1, 16]. These approaches inform community members about new resources, who is online in the community at a given time, who has joined/left the community and when. Similarly to visualization approaches, notifications provide more passive support – each member is given the same information and it is up to the user to interpret how to interpret this information. It may well be the case that members do not realise how their behaviour can contribute to the effective functioning of the community as an entity. In our approach, we identify semantic similarities between members based on their activities and aim at explicitly notify a member of how a resource made available is useful for him/her and how his/her relationship with a member can be exploited for knowledge sharing or/and collaboration. Distinctively, we aim at steering the community knowledge sharing behaviour in a way that can benefit not just individuals but the overall functioning and sustainability of VC.

Graph-based approaches for detecting interactions in community networks have been applied in different domains. Predominantly, such approaches concentrate on the structural relations of people in a graph considering the number of interactions between them [4, 18]. The key contribution of our approach is the modelling of semantic relationships via graphs, i.e. an edge connecting two members represents their semantic similarity to each other, and the relevance of this link to the community's context. This, combined with the theoretical underpinning, enables us to a develop graph-based approach for *qualitative* analysis (as opposed to structural analysis) and automatic detection of relevant interaction patterns.

The rational for detecting knowledge behaviour patters will be discussed in the following section.

3. WHAT PATTERNS SHOULD BE DISCOVERED?

This research aims to support members of a closely-knit community to answer questions like "Who knows more about subject A?", "Do others in this community know what I know?", "Who shares the most valuable resources in this community?", "Whose knowledge is important to me?", "To whom is my knowledge important?", and "What others are doing in this community?". Application of theories on TM, SMM, CCen, and collaboration support, can help to provide answers to the above questions.

Studies in CSCW community, also looked at the above awareness issues, and stress that the outcome of a member's action in the community, can influence others' actions [14]. Monitoring what others are doing and how members are related in the community is vital for: knowledge sharing, collaboration and community sustainability.

Explicitly making people aware of their similarities, in addition to activity awareness, can influence their actions and thus help them engage in the community. Consequently, discovering patterns that promote a good TM system, establishment of SMM and by exploitation of CCen, can aid at supporting the knowledge sharing in VC [8]. In this line, we are looking at four types of knowledge sharing behavior patterns as follows: (a) *two members have semantic relationships with the same members but not among themselves,* (b) *two members have semantic relationships with the same people and among themselves* (c) *members' activity in the community (uploading /downloading)* and (d) *a combination of* *the activity of a member and his/her semantic relationships with others.*

A pattern is important if can be detected, and used in order to provide support to community members. In (a) people are interested in the same areas but are not aware of the other member being interested in that area too. This is an indication of a poor TM system [17] for a VC. In addition, even if those members are related as in (b), they might not be aware of how similar they are. Making them aware of this similarity can promote collaboration among community members. Similarly, (c) can be used to identify lurkers (people who only read but do not contribute to the community) and to encourage them to contribute; or people who only upload to the community and help them realize how they can benefit from this community. Finally, (d) can identify where the CCen of a member is coming from and how to use this information to encourage knowledge sharing in the community.

The next section will introduce the main notation used to represent relationships between members in graphs and to extract the relevant patterns.

4. RELATIONSHIPS AS GRAPHS

A major part of the community model is the relationships model. What follows will provide brief definitions of each relationship and how these relationships can be detected using graph theory.

4.1 Definition of the relationships model

Assuming two community members a and b, then $ReadRes(a,b)$ relationship indicates that resources uploaded by member b are read by member a, and its strength corresponds to the relevance of the resources to the community context. $ReadSim(a,b)$ indicates that members a and b have read semantically similar resources, while $UploadSim(a,b)$ indicates that a and b have uploaded similar resources. $InterestSim(a,b)$ relationship represents the similarity of interests between members a and b. The algorithms used for extraction of the relationships model has been extensively discussed on [11] and it is out of the scope of this paper.

The relationship model is derived as graphs with edges and nodes, thus graph notation can be used to describe patterns that considered as important to detect. For each relationship we derive a graph that represents the relations between people based on that relationship type. $G_{RS}(V_{RS},E_{RS})$ represents the graph derived for $ReadSim$, $G_{US}(V_{US},E_{US})$ represents the $UploadSim$ graph, $G_{IS}(V_{IS},E_{IS})$ the graph for $InterestSim$ and $G_{RR}(V_{RR},E_{RR})$ corresponds to the graph extracted for $ReadRes$.

$ReadSim$, $UploadSim$ and $InterestSim$ are non-directed graphs of type $G(V,E)$ where V represents the node set, and E represents the edge set for each relationship type. In the edge set E the order that the vertices appear is not considered. As long as v_j and v_i are the endpoints of edge e, they are considered adjacent and are neighbors. Thus if $e \in E$ then $e=(v_i,v_j)$ and is the same as $e=(v_j,v_i)$. In the relationships network each node represents a member of the community and at each graph the nodes remain the same. An edge represents the existence of a relation between two members with strength calculated based on the algorithms presented in [11] and this differs from graph to graph. We assume that an edge is considered to be in a relationship graph ($e \in E(G)$) only if the weight of that edge is greater than the threshold set for each relationship ($w(e) > \sigma_t$). A neighborhood of a node v represents the ego network of that node and is denoted as $\mathrm{N}_G(v)$. In this paper, the neighborhood of a member a in a graph, indicates the members that a has similarity with. For example in a $ReadSim$ graph $G_{RS}(V_{RS},E_{RS})$, the neighborhood of member a, $\mathrm{N}_{RS}(v_a)$, indicates the members that a has $ReadSim$ with.

$ReadRes$ is a directed graph of type. The direction of the edge represents that a member (head) has read a resource uploaded by another member (tail). In $ReadRes$, the neighborhood of a node can be $\mathrm{N}_G^+(v):\{x \in V(G): v \to x\}$, denoting the out-neighborhood of node v and $\mathrm{N}_G^-(v):\{x \in V(G): x \to v\}$, denoting the in-neighborhood. In our context $\mathrm{N}_{RR}^+(v_a)$ represents the members who have downloaded resources uploaded by a and $\mathrm{N}_{RR}^-(v_a)$ represents the members a has downloaded resources from.

4.2 Detection of relationships in graphs

Based on the above definitions, let us denote member a and member b to be members of a VC. A relation $ReadSim(a,b)$ exists if $e_{ab}=(v_a,v_b)$ or $e_{ba}=(v_b,v_a)$ is in the set of E_{RS}. $UploadSim$ and $InterestSim$ can be detected in the same way using the respective graphs for each relationship type.

However, $ReadRes(a,b)$ exists if and only if an edge $e_{ab}=(v_a,v_b)$ and $e_{ab} \in E_{RR}$, denoting that b has read a resource uploaded by a.

In addition to the relationship graphs, information collected on the community model is also used for the automatic detection of patterns. $CCen(a)$ indicates how important is the knowledge that member a holds for the rest of the community and calculated as in [11]. $uRate(a)$ and $dRate(a)$, denote the uploading and downloading rate of member a respectively and are stored on the Individual User Model [10].

5. AUTOMATIC PATTERN DETECTION

Patterns considered as important have been discussed in section 3. In this section we will describe in detail how important patterns can be automatically detected.

P1. Unexplored similarity between community members:

For example, two members have $ReadSim$, with the same members but not among themselves.

Importance: Identifying the above situation and making people aware of their unexplored similarity with others may motivate them to participate more actively see [6]. In addition, helping them to understand that they hold complimentary knowledge improves the community TM [17] system and can promote collaboration within the community [8]. Also making them aware of what others are doing and how they relate to others promotes the building of SMM.

Detection: Let us denote members a and b to be members of the same community, thus $v_a \in V_{RS}(G_{RS})$ and $v_b \in V_{RS}(G_{RS})$. To detect this pattern, we need to extract the neighborhood of both members from the *ReadSim* graph, $G_{RS}(V_{RS}, E_{RS})$, so for member a, $N_{RS}(v_a)$ and for member b, $N_{RS}(v_b)$. If the union of their neighborhoods is a non-empty set, thus a and b have *ReadSim* with same members, and if one of the members does not belong to the other's neighborhood then P1 exists and can be described as:

$$(N_{RS}(v_a) \cap N_{RS}(v_b) \neq \emptyset) \wedge (v_a \notin N_{RS}(v_b))$$

The same pattern can be detected for *UploadSim* and *InterestSim* in the same way.

P2. Community members may not be aware of their similarity:

Two members have *ReadSim*, with the same members and among themselves.

Importance: Community members are not aware of how similar they are in terms of uploading, reading or interests with other members of the community. Detection of this pattern can be used to promote building of SMM [13], (knowing what others are doing), and promoting TM [17], (who holds similar knowledge as I am?), in the VC.

Detection: This pattern can be detected by extracting the neighborhood of both members from the *ReadSim* graph, $G_{RS}(V_{RS}, E_{RS})$, similarly as for P1. If the union of their neighborhoods is a non-empty set, thus a and b have *ReadSim* with same members, and if one of the members belongs to the other's neighborhood then P2 has been identified and can be described as:

$$(N_{RS}(v_a) \cap N_{RS}(v_b) \neq \emptyset) \wedge (v_a \in N_{RS}(v_b))$$

The same state can be detected for *UploadSim* and *InterestSim* in the same way.

P3. Problems with participation:

A member appears to upload resources but not download.

Importance: Participation to the community is important to be monitored for each member. This pattern can be useful to identify members who are not downloading from the community. Support can be provided to those members in order to benefit and make the most of their time in the community.

Detection: Detection of P3 is done by using the upload and download rate of a member. P3 exists if member a has uploading rate, ($uRate(a)$), greater than 0 and downloading rate, ($dRate(a)$), equal to 0. This can be expressed as:

$$(uRate(a) > 0) \wedge (dRate(a) = 0)$$

P4. Inactive members:

A member who appears to *download but not upload* resources to the community can be detected similar to P3 and can be denoted as:

$$(uRate(a) = 0) \wedge (dRate(a) > 0)$$

P5. Important peripheral members not downloading:

A member who appears to uploading but not downloading and others are reading what he/she is uploading.

Importance: In this case, we can use the information to motivate that member to start benefiting from the community. By showing to him/her that others are interested to what he/she is uploading, it can motivate that member to start reading resources uploaded by the members he/she has similarity with.

Detection: The first part of P5 is calculated similarly to P3 by checking the uploading and downloading rates for member a. Then from the *ReadRes* graph, $G_{RR}(V_{RR}, E_{RR})$ we have to extract the out-neighborhood of member a, ($N^+_{RR}(v_a)$), and check if it is non-empty, indicating that others are reading what member a is uploading. Having done this we can extract P5 using the following:

$$(uRate(a) > 0) \wedge (dRate(a) = 0) \wedge (N^+_{RR}(v_a) \neq \emptyset)$$

P6. Important peripheral members not uploading:

A member appears to download only and has *InterestSim* with other members.

Importance: This pattern can be used to motivate people who are only downloading from the community to start uploading, by showing them how similar, in terms of interests, they are with other members. This kind of awareness can improve the TM system of the community since members will be aware of others' interests [17]. By motivating them to upload to the community improves the community to sustain.

Detection: To detect P6 we need to check that the uploading rate for a member is zero, thus this member is not uploading and the download rate is greater than zero, thus this member is downloading from the community. The *InterestSim* neighborhood, ($N_{IS}(v_a)$), has to be extracted for member a from the *InterestSim* graph $G_{IS}(V_{IS}, E_{IS})$ and check to be non-empty. The formal description can be found below:

$$(uRate(a) = 0) \wedge (dRate(a) > 0) \wedge (N_{IS}(v_a) \neq \emptyset)$$

P7. Unexplored complimentary similarity between members:

A member appears to have *UploadSim* with a given member but he/she does not have *ReadSim* with that member.

Importance: Members who upload similar resources in the community but are not reading similar resources, have similar and complimentary interests. We need to make these people aware of their similarity in terms of uploading and pinpoint their difference in terms of reading. This intervention will improve the TM system of the VC since people will be able to identify where important knowledge, for them, is located [7], and at the same time improve the building of SMM [13], since they will start appreciating others activity in the community. Also by telling them about their difference in reading resources we are promoting awareness of where complimentary knowledge is located and thus encourage collaboration.

Detection: P7 can be identified using the *UploadSim* graph $G_{US}(V_{US}, E_{US})$, and the *ReadSim* graph $G_{RS}(V_{RS}, E_{RS})$. Assuming members a and b are members of the same community and v_a, v_b represent those members in the above graphs. We need to check if one of the members appears on the other member's neighborhood for in the G_{US} and not appearing on the neighborhood of that member's G_{RS} , thus:

$$(v_a \in N_{US}(v_b)) \wedge (v_a \notin N_{RS}(v_b))$$

The above patterns have been developed and applied to tracking data extracted from a real closely-knit virtual community. The next section describes the application of the above algorithms, and gives a detailed report on what has been discovered and how it affects, the development of TM, establishment of SMM and act as barrier to collaboration, in the virtual community.

Table 1 gives a summary of the importance of the detection of each pattern to the processes we are considering.

Pattern	Affects
P1	Collaboration, TM System, SMM
P2	SMM, TM System
P3	Sustainability
P4	
P5	SMM, Collaboration
P6	TM System, Collaboration
P7	SMM, TM System, Collaboration

Table 1. Summary of how the detection of a pattern can affect the relevant processes

6. STUDY WITH A BSCW VIRTUAL COMMUNITY

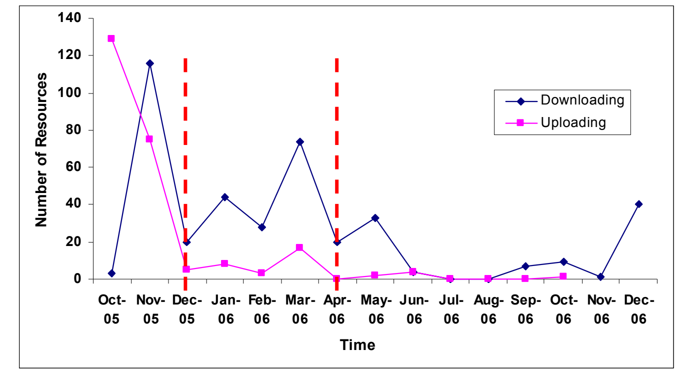
To validate the knowledge sharing behaviour pattern algorithms we have employed them to extract patterns of a real community which both authors belonged to. The VC in our study included 34 members (researchers and doctoral students) from two research groups working on similar research areas, sharing documents and research papers with the BSCW system that provides general support for collaboration over the web [2]. The groups were based in two European countries, some members knew each other but many had never met. The community was established in 2003.

We collected log data from October 2005 until December 2006 using BSCW features allowing every member to see what is happening in the community.

The algorithms extracted relationships between 1122 pairs of members, (34 members x 33 members), this is the relationship between any two members apart themselves. There was a gradual decline in the uploading and downloading of resources in the observed period. During the beginning of the monitored period (October 2005 – May 2006) members were uploading and downloading papers. After that the activities minimised for all members, and during the last few months of the monitored period (October 2006 – December 2006) there was no uploading and very little downloading. The community was gradually declining and has almost stopped its activity at the moment. The study was conducted to identify problems that could have been spotted earlier and addressed properly to help this community sustain. Consequently, the data collected have been divided in such a way to aid identify where interventions could have been injected. At first, pattern algorithms have been applied on the data collected from October 2005 – May 2006 inclusive. After May 2006, members began to loose interest to the community and minimise their activity. Although results showed that problems could have been detected we needed more specific indications on when actually interventions could have triggered. Having a look at the community activity (fig 1) we can identify two points where the activity minimised rapidly (December 2006 & April 2006).

Thus, we applied the algorithms on the data collected in January 2006, February 2006 and March 2006, the three months after and just before the two activity drop occasions. This further dissection of data will give us the indication on where interventions could have been done in order to support the sustainability of the community.

Figure 1. Uploading (blue) and downloading (pink) activity in the BSCW virtual community. The red dotted lines show the drop of activity in two different occasions.



The next section will discuss the results obtained after the algorithm application.

7. APPLICATION OF PATTERN ALGORITHMS

The log data was stored in a text file, fully anonymised, and then converted to database tables. The tables were used as input for the relationship modelling algorithms presented in [11] and implemented in Java. Relationships and activity data were extracted as MySQL database tables and used as input to the pattern algorithms that implemented also in Java. This section will give a brief overview of the algorithm application results on data collected from October 2005 – May 2006 and will focus on the results obtained by the application of the algorithms on the data collected on January 2006, February 2006 and March 2006. We will show here representative examples of patterns discovered, and will discuss how this can be used for interventions.

7.1 Results October 2005 - May 2006

The application of the patterns on the data collected between October 2005 to May 2006 uncovered that the community suffered from lack of TM system, and SMM and thus collaboration between community members was difficult to be achieved.

Detection: P1 reveals 195 pairs of members to have a relationship with the same members in the community but not among themselves.

Importance: This affects the TM system of the community since members are not aware of who are similar to in the community. Not knowing who is working on similar projects or is interested on similar areas as you, is a barrier to collaboration.

Detection: 135 pairs of members in the community have a relationship with the same people and among themselves, based on the results from P2.

Importance: Are they aware of this relationship they have? Do they know how similar they are to other people? The development of SMM and TM system can be affected if people do not know what others are working on in the community.

Factors that usually affect the lifetime of a VC are members who are only benefiting from, or those who are only uploading to, the community.

Detection: In the VC under study, 2 people appear to only uploading during the study period and 13 were only downloading (P3 and P4).

Importance: These phenomena affect the sustainability of the community since are breaking the chain of knowledge sharing. Having so many people only downloading from the community limits the new material that is coming in, leading to other members loosing interest to the VC and become inactive, causing the community to stop functioning.

In order to support the community to sustain and improve the knowledge sharing practice a deeper analysis on the data collected in January 2006, February 2006 and March 2006, (well before the community stop functioning), had to be done. The results provide us with data that can be used to provide intelligent support to the VC.

7.2 Results January, February & March 2006

The results show that each month the community was coming closer to stop functioning. More patterns, thus problems, have been identified in March 2006 than in January 2006. This section will provide analysis of the results obtained and how these relate to the patterns described in section 5.

P1. Unexplored similarity between community members:

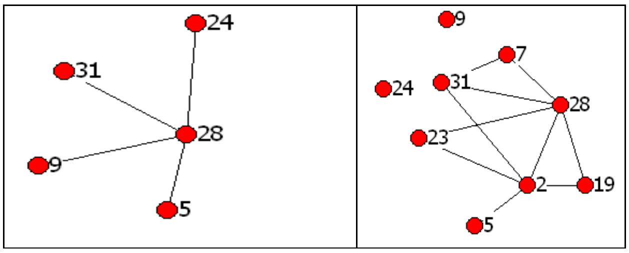
In January 2006 we had ten pairs of members who had relationship similarities with the same members but not among themselves. In March twenty-two pairs of members find themselves in that situation.

Importance: This situation creates a problem to the VC as it shows that a TM system is not in place, people are not aware of whom they have similarity with and there is lack of SMM since members do not know what others are working on.

Detection: In January 2006 members M9, M24, M31, and M5 appeared to have *ReadSim* with M28 but not having *ReadSim* among themselves. M28 appears to be a connecting node between these four members (Figure 2 Left). In February 2006 members M5 and M31 continue to find themselves in the same situation but members M9 and M24 stop contributing or downloading from the community (Figure 2 Right).

Support & Benefits: Automatic messages could have been sent to members M9, M24, M31 and M5 to make them aware of their similarity through M28 and consequently to be kept motivated to contribute. This would have add to the development of a better TM system and open the doors for possible collaboration between those members. This detection could have been used to make M28 aware of his/her important role in the community and encourage him/her to begin collaborating with M5, M9, M24, and M31 and also to keep him/her motivated. This intervention could have been used to keep members M9 and M24 active in the community, contribute to the building of SMM and facilitate knowledge sharing.

Figure2. Left: ReadSim in January 2006 showing M9, M24, M31 to have ReadSim with the same members as M5 but not with M5. Right: ReadSim in February 2006 showing members M9 and M24 to be disengaged



Detection:In January 2006, members M5 and M9 appear to have *InterestSim* with M24 and M28 but not among themselves. Similarly M23 has *InterestSim* with 28 (Figure 3). This shows that members M5, M9, M23, M24, and M28 have closely related interests but might not be aware of their similarity with each other. In February and March members M9, M23, and M24 became inactive.

Support & Benefits: Support in the form of automatic messages could have been provided to those members to realize their interest similarity with other members in the community and thus motivate them to remain active.

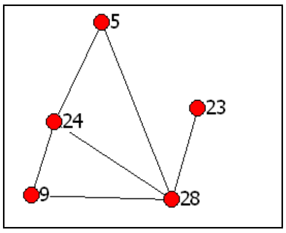


Figure 3. In January 2006 members appearing to have InterestSim with the same members but not among themselves.

P2. Community members may not be aware of their similarity:

Importance: Members who have a relationship with the same people and among themselves have to also be supported. We need to ensure that these members are aware of the relationship they have and its importance to them and others in the community. This will ensure that members establish SMM since they will be aware of what others are working on and promote the building of TM system since they will know who they relate to in the community and how similar they are to others.

Detection: Good examples for this pattern are members M9, M24 and M31. Members M9 and M24 in January, appear to have *ReadSim* among themselves and with member M31 (Figure 4 Left). In February, M9 and M24 disengage from the community. Where these members aware of the similarity they had with each other and member M31? Additionally, in February members M2, M5, M7 and M19 appear to have *InterestSim* with each other (Figure 4 Right). In March members M2 and M7 disengaged from the community and members M5 and M19 minimize their activity.

Support & Benefits: We need to support these members and make them aware of the *ReadSim* and *InterestSim* they have with each other. This would encourage them to explore the links they have and improve the SMM of the community and also contribute to a better TM system.

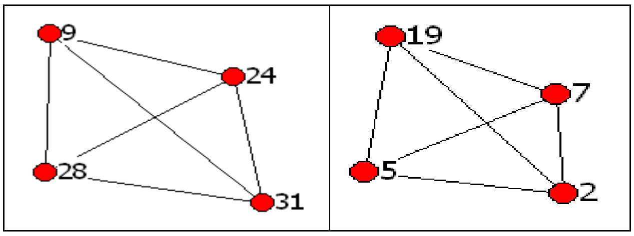


Figure 4. Left: ReadSim between members in January 2006. Right: InterestSim between members in February 2006.In both cases members appear to have similarity among themselves but are they aware of the similarity they have?

P3 & P4. Problems with participation and Inactive members:

As observed on the results from October 2005 – May 2006, there were members who only download and members who only upload in the community.

Detection: In January 2006 one member was only uploading and four members were only downloading. By March 2006 two members were only uploading and nine members only downloading.

Importance: The increase of people identified in March 2006 shows that members began to disengage from the community either because they lost interest or because they cannot find information useful for them.

Detection: Only downloading excessively is a behavior that newcomers develop when they struggle locate information important to them. For example in March 2006 member M19 downloaded thirty-three resources, without uploading anything from January to March.

Support & Benefits: Members like M19 can be supported by providing him/her information of members with similar interests based on his *InterestSim* or members who are reading similar resources (*ReadSim*) or uploading resources similar to M19's interests (*UploadSim*). This will help that member position him/her self in the community, motivate him/her to contribute and at the same time improve SMM, the development of TM system and facilitate knowledge sharing.

P5. Important peripheral members not downloading:

Members, who sit in the periphery of the community, uploading interesting resources but do not benefit from the knowledge available in the VC, need to be supported.

Importance: We need members to engage in the community by both uploading and downloading in order to keep the knowledge sharing practice active. Furthermore with the detection of this pattern we are promoting collaboration since members are becoming aware of people who are reading what these members are uploading. With the awareness we offer, we are also building SMM in the VC.

Detection: Furthermore, member M33 detected to uploading but not downloading from the community and other members to be interested on what M33 is uploading. But this person disengaged from the community shortly after March 2006.

Support & Benefits: This information could have been used in order to motivate members who are detected with this behavior to stay in the community and benefit from people with similar interests. Member M31 who is a cognitively central member, used to read resources uploaded by member M33. If M33 knew about the *ReadRes* he/she had with M31, he/she could have become interested on the resources uploaded by M31 and remain active in the community.

P6. Important peripheral members not uploading:

Several members appear to only download resources and they have *InterestSim* with other members in the VC.

Importance: Similarly to P5 members need to contribute to the VC as well as benefiting from what is available. This pattern can be used to identify members who are not uploading in the community and have *InterestSim* with others. This detection can be used to promote collaboration and improve the TM system of the community since people will become aware of others' interests.

Detection: A good a example for this pattern is member M5. In January 2006 members M5 appears to download only and to have *InterestSim* with members M24 and M28. In February 2006 M5 appeared to still only downloading and have *InterestSim* with members M2, M7 and M19.

Support & Benefits: The above detection can indicate that M5 needs support to identify people in the community or resources of his/her interest. Downloading excessively can indicate that M5 is confused and doesn't know how to benefit from the community. The similarity he/she has with others can be used to send automatic messages as a motivation for M5 to start contributing since there are members with similar interests as his/hers. This will promote the collaboration among members with similar interests and make M5 aware of others' knowledge, thus promote TM system.

P7. Unexplored complimentary similarity between members:

Importance: It is important for members to identify not only members who might hold similar knowledge as they do but also members who might hold complimentary knowledge to theirs. This allows members to identify potential collaborators, it is promoting the building of TM system since they become aware of others' knowledge and through the awareness it promoted SMM.

Detection: Member M13 appears to have *UploadSim* with member M33 but not have *ReadSim* with that member.

Support & Benefits: This information could have been used in order to keep both M33 and M13 aware of their similarity and motivate them to read resources uploaded by each other. This would result in possible collaboration.

Through the graph based detection algorithms we are able to automatically identify problems inside the community and use those detections to provide support to the community as a whole. Members who have a similarity in interests or in the resources they are reading, and are not aware of it, can now be supported and become aware how they relate to each member in the community. Also members who remain to the periphery of the community can be supported to identify where knowledge or people important to them are located in the VC in order for them to get motivated and move to the centre of the VC. Similarly members who are not uploading can be detected and through interventions can be motivated to become active, for the community to sustain. Although not every situation discovered has been discussed in this section, the examples used are representative of what is detected and how the detection can be used.

The patterns described above correspond to the knowledge sharing behavior patters identified at [11]. Compared to the method used in [11], the approach described here allows us to identify patterns in a data of larger scale and use those patterns directly as input for interventions, in order to support the community members. Additionally, our approach proved successful in identifying where and when support is needed in order to help community members have a good experience while the community is active.

8. CONCLUSION

In this paper we proposed a new approach of identifying knowledge sharing behavior patters in a VC and use those in order to provide community-tailored support. The study performed revealed convincing results and allowed us to evaluate the graph based algorithms developed by identifying when interventions could have been done to support community members.

The next step of this research project will be to apply the methods described in this paper and in [11] in an active VC in order for the whole framework to be evaluated along with the effectiveness of the automatic interventions.

9. REFERENCES

1 Agostini, A., et al. *Stimulating knowledge discovery and sharing*. in *GROUP '03: Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*. 2003: ACM.

2 Applet, W. *WWW Based Collaboration with the BCSW System*. in *SOFSEM'99*. 1999. Milovy, Czech Republic: Springer Lecture Notes in Computer Science 1725.

3 Barley, S., et al. *Does CSCW need organization theory?* in *CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work*. 2004. Chicago, Illinois, USA: ACM.

4 Falkowski, T. and M. Spiliopoulou, *Users in Volatile Communities: Studying Active Participation and Community Evolution*, in *User Modeling*. 2007. p. 47-56.

5 Fischer, G. and J. Ostwald, *Knowledge Management: Problems, Promises, Realities, and Challenges.* IEEE Intelligent Systems, 2001. **16**(1): p. 60 - 72.

6 Harper, M., et al. *Talk amongst yourselves: inviting users to participate in online conversations*. in *IUI 2007*: ACM.

7 Hollingshead, A.B. and D.P. Brandon, *Potential Benefits of Communication in Transactive Memory systems.* Human Communication Research, 2003. **29**(4): p. 607 - 615.

8 Ilgen, D.R., et al., *Teams in Organizations: From Input - Process - Output Models to IMOI Models*, in *Annual Review of Psychology* 2005. p. 517 - 543.

9 Kay, J., et al., *The Big Five and Visualisations of Team Work Activity*, in *Intelligent Tutoring Systems*. 2006. p. 197-206.

10 Kleanthous, S., *Semantic-Enhanced Personalised Support for Knowledge Sharing in Virtual Communities*, in *User Modeling 2007*. 2007. p. 465-469.

11 Kleanthous, S. and V. Dimitrova, *Modelling Semantic Relationships and Centrality to Facilitate Community Knowledge Sharing*, in *Adaptive Hypermedia & Adaptive Web-Based Systems (AH'08)* 2008: Germany. p. 10.

12 Lave, J. and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*. 1991, New York: Cambridge University Press.

13 Mohammed, S. and B.C. Dumville, *Team mental models in a team knowledge framework: expanding theory and measurement across disciplinary boundaries.* Journal of Organizational Behavior, 2001. **22**(2): p. 89 - 106.

14 Schmidt, K., *The Problem with `Awareness': Introductory Remarks on `Awareness in CSCW'.* Computer Supported Cooperative Work (CSCW), 2002. **11**(3 - 4): p. 285-298.

15 Shami, S., et al. *That's what friends are for: facilitating 'who knows what' across group boundaries*. in *GROUP '07: Proceedings of the 2007 international ACM conference on Supporting group work*. 2007: ACM.

16 Wang, Y., W. Graether, and W. Prinz. *Suitable notification intensity: the dynamic awareness system*. in *Proc. of the int. ACM conf. on Supporting Group Work*. 2007.

17 Wegner, D.M., *Transactive Memory: A Contemporary Analysis of the Group Mind*, in *Theories of Group Behavior*, B. Mullen and G.R. Goethals, Editors. 1986, Springer-Verlag. p. 185 - 208.

18 Zhang, J., M. Ackerman, and L. Adamic. *Expertise networks in online communities: structure and algorithms*. in *WWW '07: Proceedings of the 16th international conference on World Wide Web*. 2007: ACM.