Modelling Semantic Relationships and Centrality to Facilitate Community Knowledge Sharing

Styliani Kleanthous and Vania Dimitrova
School of Computing, University of Leeds
{stellak,vania}@comp.leeds.ac.uk

Abstract. Some of today's most widely spread applications are social systems where people can form communities and share knowledge. However, knowledge sharing is not always effective and communities often do not sustain. Can user modelling approaches help to identify what support could be offered and how this would benefit the community? The paper presents algorithms for extracting a model of a closely-knit virtual community following processes identified as important for effective communities. The algorithms are applied to get an inside of a real virtual community and to identify what support may be needed to help the community function better as an entity.

Keywords: Knowledge Sharing, Community Model, Community Adaptation

1 Introduction

Social systems, which enable people to form communities and share knowledge, are becoming increasingly popular nowadays. Studies have shown that having technology and people present does not guarantee the sustainability of a virtual community (VC) [4]. Appropriate support is needed to facilitate the functioning of a community where the members actively engage and share knowledge effectively [6]. Along this line, personalisation and adaptation can play a crucial role, as illustrated by recent user-modelling approaches [2, 13]. However, existing adaptation techniques focus mainly on supporting individual members, rather than supporting the community to function *as an entity*. We propose a new method for community-tailored support which is aimed at facilitating processes related to the effectiveness and sustainability of VCs [9] and is based on a community model derived from analysis of log data.

In a broad sense, virtual communities vary from fairly large, loosely structured communities to relatively small, closely-knit ones. In this paper we consider closely-knit VCs for knowledge sharing, which are characterised by a common purpose, participants' commitment to the sharing of information and generation of new knowledge, and equal membership inside the community. Closely-knit VCs usually exist in relatively well-defined organisational or educational settings, and can share common characteristics with teams. Following research in organisational psychology [7], we have identified several processes important for effective team functioning which can be applied to VCs and can be examined or facilitated by analysing community log data. These processes include: good *transactive memory system* (members are aware how their knowledge relates to the knowledge of the others), *shared mental models* (members develop a shared understanding of the key processes and the relationships that occur between them), and *cognitive centrality* (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 [11]).

Based on the above processes, we have defined algorithms to extract a *community model* that includes: individual user models of all members, a model of the semantic relationships in the community, a list of the cognitively central members, a list of the popular and peripheral topics in the community, and the community context defined by the key topics of interest within the community [10]. This paper will present the application of the community modelling algorithms to get an understanding of what is happening in a real community, and to identify what support can be provided to improve the functioning of this community. We will focus on semantic relationships and cognitive centrality, which are the kernel of the community model in our approach, see [10] for a description of all components. Section 2 describes the algorithms developed to extract the community model. Section 3 presents the study performed using a VC and Section 4 discusses application of the community model and points out possibilities for community tailored support. Related work is discussed in Section 5, and the last section points at future work.

2 Semantic Relationships and Cognitive Centrality

This section will outline the main algorithms to capture the semantic relationships and centrality within a community. As input we consider the *metadata of the resources* shared in the community, such as: (a) keywords associated with each resource ($rKeywords$), which can be provided by the publisher or by the members in terms of tags, (b) the person who shared or accessed a specific resource, and (c) the time when a resource is uploaded or read. We also consider the *community context* C_T which is the list of key topics for this community. A formal description of the input is given in [10]. We will consider four types of semantic relationships between users: $ReadRes$ relationship indicates links based on reading resources uploaded by others, $ReadSim$ and $UploadSim$ indicate relationships based on similarity of read or uploaded resources, respectively, and $InterestSim$ indicates similarity in members' interests. We combine these relationships to calculate the cognitive centrality of each member.

2.1 ReadRes Relationship

$ReadRes(a,b)$ 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. $ReadRes$ can be used to identify complementary knowledge among people, and this helps to improve the community's *transactive memory*[15].

Consider a resource r_i uploaded by b and read by a. We will denote its keywords with $rKeywords_i$. Considering the community context C_T, we define the value of r_i for the community as $V_{r_i} = Sim(rKeywords_i, C_T)$. To calculate the similarity between two lists of terms we employ a similarity algorithm based on WordNet [12]

Let us denote $N_r^{a \leftarrow b}$ to be the number of resources uploaded by b and read by a. The value of $ReadRes(a,b)$ is the sum of all values of the resources uploaded by b and read by a, based on their relevance to the community context, i.e.:

$$ReadRes(a,b) = \sum_{i=1}^{N_r^{a \leftarrow b}} V_{r_i}$$

2.2 ReadSim and UploadSim Relationships

$ReadSim(a,b)$ indicates that members a and b have read semantically similar resources, while $UploadSim(a,b)$ indicates that a and b have uploaded similar resources. These relationships can be important for discovering similarities that members may not know about. By making people aware of who else is holding knowledge similar to theirs can improve the community's *transactive memory system* [15]. This can also improve the understanding of what is happening in the community which can be related to the development of *shared mental models[7]*. *UploadSim* can also be used to identify people who are not uploading and to encourage them to contribute by pointing at their *ReadRes* or *ReadSim* relationships with others.

To calculate $ReadSim(a,b)$ we derive an extended list of keywords for each member by combining the keywords of every resource read by this member. Let us denote these extended keyword lists as $aKeywords$ and $bKeywords$. These lists are compared to find the similarity between them by using again the WordNet similarity algorithm described in [12]. Hence, the $ReadSim(a,b)$ is calculated as follows:

$$ReadSim(a,b) = Sim(aKeywords, bKeywords)$$

$UploadSim(a,b)$ is calculated similarly using the resources uploaded by a and b

2.3 InterestSim Relationship

$InterestSim(a,b)$ relationship represents the similarity of interests between members a and b. This relationship can identify interest complementarities. Furthermore, making members aware how their interests relate to the others can motivate participation. Finding people with similar interests and making them aware of this similarity can indicate possibilities for *collaboration*. Awareness of other people's interests can improve the shared understanding the members have about the community and help the development of *shared mental models* [7].

To derive interests of a member, we considered the resources he/she has uploaded and downloaded. Using the keywords $rKeywords$ for each resource uploaded or downloaded by a user, his/her interests are represented as a list of terms with weights.

For example, all terms that member a has shown any interest in are aggregated in the list T_a, where every term $t \in T_a$ has weight $w(t,T_a)$ that indicates the frequency of t in T_a. If $w(t,T_a) \geq \sigma$ (σ is a threshold), t is added to the interests of a denoted with I_a,

I_a is presented as the member a's personal list of interests. The same algorithm is used to derive the personal list of interests for member b - I_b. I_a and I_b are compared with the algorithm in [12] to calculate the interest similarity between a and b:

$$InterestSim(a,b) = Sim(\mathrm{I}_a,\mathrm{I}_b)$$

2.4 Cognitive Centrality (CCen)

Cognitive centrality measure is used to locate knowledge inside the community that is important to the community members. This can be helpful to identify the central members and how they contribute to the community. It can also be useful in identifying unique knowledge held by peripheral members. This is important for the community's sustainability and flexibility - interests might shift in time [11], knowing where unique knowledge is located can facilitate the transition from one subject area to another [15]. Being aware of the central and peripheral members of the community can also help the improvement of *shared mental models* and *transactive memory.*

To calculate each member's centrality within the community, we adapt the degree centrality algorithm used in social networks [5]. $CCen(a)$ of member a is calculated as the number of all members b to whom a is connected considering the four relationship types defined above:

$$CCen(a) = \sum_{b=1}^{n} ReadRes(a,b)+ReadSim(a,b)+UploadSim(a,b)+InterestSim(a,b)$$

The above algorithms were applied to extract a community model based on tracking data from an existing closely-knit virtual community. The next section describes a study that shows how the community model derived can be used to determine what support may be offered to improve the functioning of the community.

3 Study with a Virtual Community

To validate the community modelling algorithms we have employed them to extract a model 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 [14]. 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 BSWC features allowing every member to see what is happening in the community.

The activity monitored included uploading and downloading resources, 244 resources in total. Four members were only uploading while thirteen were only downloading. Eight members were isolates and never uploaded or downloaded resources. Figure 1 shows the gradual decline in community contribution in the observed period. Downloading was also declining - as shown in figure 2, most members reduced their downloading activity except for members 2 and 3.

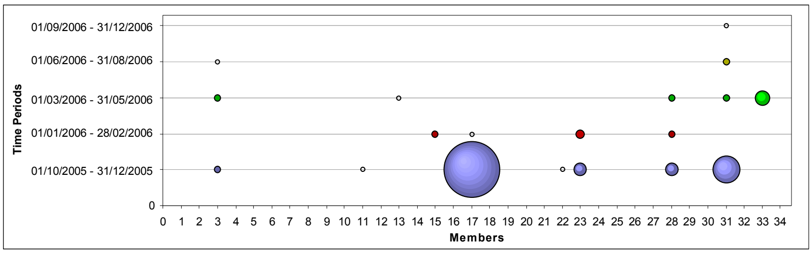

Fig. 1. Uploading: the ball size corresponds to the amount of resources uploaded by a member.

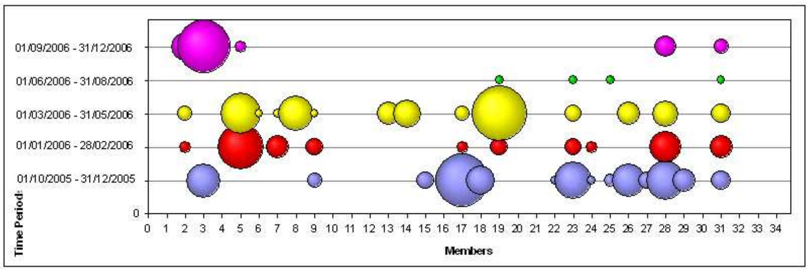

Fig. 2. Downloading: the ball size corresponds to the resources downloaded by a member.

4 Application of the Community Model

The community log data was stored in a text file, fully anonymised, and then converted to data base tables. The tables were used as input for the community modelling algorithms described in Section 2, and in [10], and implemented in Java. We will show here representative examples of phenomena discovered about the community, and will discuss how this can be used for adaptive support. In the illustrations below, excerpts from the community model are rendered with NetDraw[1].

4.1 Application of the Relationships Model

The relationship model indicated strong semantic relationships between community members which were often not explored by community members.

According to the community model, the members who never uploaded resources in the community had in fact *ReadRes* similarity (see figure 3). There are links with two groups – the group including members 31, 29, 15, 3, 13, 22, 23, 29, and 17 (shown in blue) and with the group of members 33, 20, and 12 (shown in black).

[1] Network visualization software: http://www.analytictech.com/Netdraw/netdraw.htm

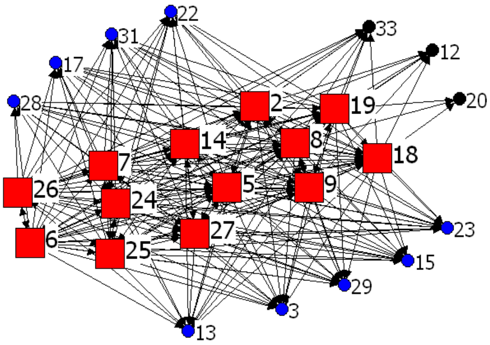

Fig. 3. The members not uploading to the community, in rectangles, have unexplored *ReadRes* similarity with the same people.

The situation in figure 3 indicates that the community's transactive memory system is not well-developed, which points at the need for appropriate support. For example, automatic messages can be generated to point out to member 29 (who is actively engaged in the community) that he/she has a relation with member 19 (who is not uploading). Providing such awareness can improve the transactive memory, develop members' understanding of what the others are doing and facilitate collaboration.

Another interesting case concerns members 7, 24 and 26 who have *ReadSim* relation with almost the same people but have no connection among themselves (figure 4). Interestingly, these people are coming from the same research group and, as indicated in the community model, they have not explored (and are perhaps unaware of) their connections via the community. Making these members aware of their similarities with others may motivate them to better participate in the community, see [6]. It can also facilitate knowledge sharing between these people, who appear to be interested in the same topics [4], and may promote collaboration.

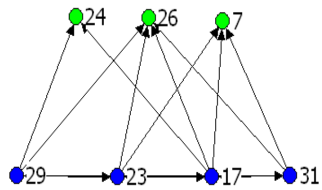

Fig. 4. *ReadSim* between members who are not aware of the relationship between them.

Most nodes of the graphs representing *ReadSim*, *UploadSim*, and *InterestSim* relationships appeared strongly connected. This confirmed our expectations for the community model (when people are working in similar areas their interests and the resources shared tend to be semantically similar). However, it also pointed out that further fine tuning of the similarity algorithms could be beneficial. For this, we are currently integrating an enhanced model of the community context represented with an external ontology and a taxonomy of resource folders, as described in [10].

4.2 Application of Cognitive Centrality

The centrality of each member (figure 5) was calculated based on the formula presented in Section 2.4. Members 31, 29 and 17 are indicated as the three most central members of this community. This closely corresponds to the real world - members 17 and 31 are the leaders of the two research groups involved in this community, while member 29 is a researcher who actively contributed to the VC.

Centrality can be influenced by different circumstances. For example, members 6 (newcomer) and 25 (oldtimer) gained some centrality due to actively downloading from the community. Such members might be aware of the cognitive processes in the community and can provide valuable information to the others. Member 13 on the other hand, is an old-timer actively engaged by both reading and uploading resources to the community. This member is indeed involved in most projects and can be quite influential to the community.

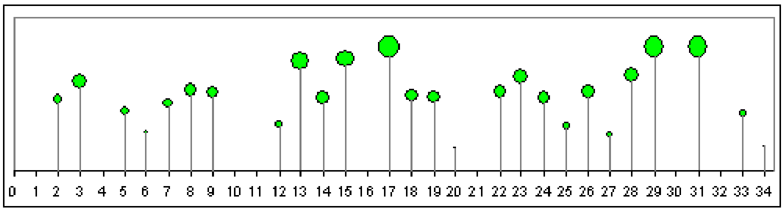

Fig. 5. Community members' centrality.

It is interesting to compare the centrality of two members 5 (newcomer) and 26 (old-timer) – who have not uploaded resources. Member 26 appears to be more central to the community than member 5 although 26 has read fewer resources (twenty-one in total) than 5 (who read fifty resources in total). This indicates that 26 has read resources that are closer to the community's interests and illustrates the effect of the community's context on deriving relationship values (Section 2).

The centrality measure can be a way to motivate people to contribute and remain active, e.g. in [2] centrality is visualised to encourage participation. We consider push algorithms where tailored messages can be sent to members based on their centrality. For example, members 5, 6 and 25 can be encouraged to contribute to the community since they already have similarities with the rest of the community. Indicating the most central members can be beneficial for the community. They can be asked to point others at valuable resources, e.g. when member 28 (peripheral) is searching for a topic which member 31 (central) seems to have information about, we can display a message to direct 28 to 31 for further help. Also, a newcomer like 6 can be integrated faster if they are mentored by a cognitively central member with similar interests.

4.3 Interesting Individual Cases

Information about individual users' engagement can be combined with the relationships model to identify cases where individuals can be given support in order to improve the functioning of the community as a whole.

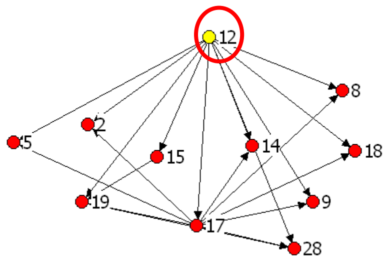

Fig. 6. *ReadRes* relationships of member 12. The graph shows the members who have read

For instance, member 12, a professor coordinating projects involving community members, has not downloaded anything, and has uploaded only one resource read by many members (figure 6). 12 was identified as a fairly central member, as what they shared was important to the community. This member can be informed that people are interested in his/her resources and that there are other members uploading similar resources. This can motivate 12 to engage and can improve the knowledge sharing.

the resource uploaded by 12.

A typical problem for the effective functioning of communities is the *integration of newcomers* (newly joining members). There were several newcomers who did not integrate in the community during the analysed period. For example, member 14 was very active during the first two months after his/her joining but then became fully disengaged. The relationships model indicates that 14 has read resources similar to those read by others and has similar interests to other members (figure 7). The community model helped recognising similar behaviour followed by other members (e.g. 25 and 19) - downloading actively for some time and becoming disengaged afterwards. This might be an indication that these members are struggling to find their way in the community's knowledge space and are uncertain about their role in the community. Such members can be helped to become aware of their cognitive relationships with the others, so they may be motivated to remain actively involved.

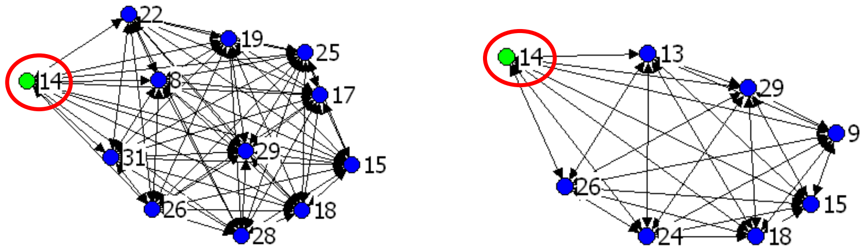

Fig. 7. *ReadSim* (left) and *InterestSim* (right) ego networks for member 14. The above networks show that the resources member 14 was reading were similar to the resources several members on the community were reading too. Also the derived interests of member 14 are similar to the interests of other community members.

Another interesting newcomer case is member 33 who was inactive at the beginning but then started contributing to the community. He/she uploaded a total of eleven resources but only one resource was read by one other member (figure 8). Member 33 was a visitor for a year at one of the research groups whose leader was member 31. The relationship model indicated that many members uploaded similar resources to 33. Unfortunately, these links were never exploited and the VC as a whole did not benefit from the knowledge "shared" by 33.

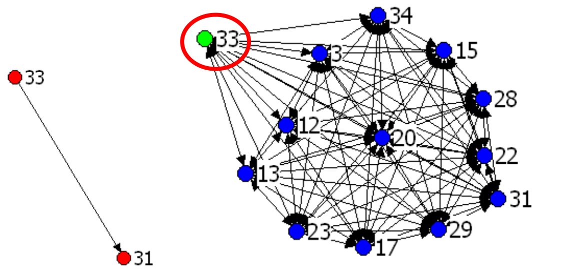

Fig. 8. *ReadRes* for member 33 (left) and ego networks for *UploadSim*(33) (right). Despite the similarity, 33 did not integrate in the community.

The example shows how the community model helped detecting an isolated niche which hinders the effective knowledge sharing. Based on *ReadSim* or *InterestSim* relationships, oldtimers that have similar interests or are reading similar resources and are actively engaged in the community can be approached.

For instance, a message can be sent to member 31 to help the newcomer 33 to integrate in the VC. Member 33 could also be reminded that others have similar interest and are uploading relevant resources. At the same time, oldtimers can be encouraged to look into interesting resources uploaded by newcomers. In general, such support aims at improving the community's transactive memory and can motivate members to remain engaged.

5 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 VCs [13]. Our approach does not aim at identifying expertise alone, but also derives a person's influence in the VC based on the relationships he/she has developed with others, which benefits the VC as a whole.

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 [8] or to show the status (or popularity) of a resource [14]. 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.

Recently research on modelling communities employed graph theory to model relationships between members [8] or members' interactions in general [3]. The key contribution of our approach to community modelling is the considering of semantic relationships, 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.

The relationship model in [1] is the closest to ours but there is a crucial difference. Users' interests are modelled in [1] based on how frequently and how recently users have searched for a specific area from the ACM taxonomy, and user relationships are derived based on any successful download or service that took place between two users. In contrast, our approach employs the metadata of the resources shared in the community and derives a semantically relevant list of interests for every user.

6 Conclusion and Future Work

We have proposed a new approach for modelling relationships and centrality in a virtual community, aimed at supporting processes that facilitate the effective knowledge sharing and sustainability of VCs. The community modelling algorithms have been employed to derive a model of a real VC, which has indicated when and how community-tailored support can be offered.

The goal of this research is to develop computational means to provide community-tailored support for knowledge sharing. We are currently tuning the community modelling algorithms by integrating an existing ontology to represent the community context. The BSCW team kindly provided us with anonymised tracking data from another community which we can use as a second case study. We plan to examine how our approach can improve awareness by comparing it to recent visualisation techniques extending BSCW [14]. Possibilities for future work include also the use of data from a different VC, e.g. Comtella [2], to further evaluate the extracted community model in real settings. Our future work will also include developing algorithms that automatically detect changes in the behaviour of a closely-knit VC, which will help us examine the possible effect of adaptive support offered.

References

1. Bretzke, H. and J. Vassileva. *Motivating Cooperation on Peer to Peer Networks*. in *9th Int. Conf. on User Modelling*. 2003. USA.
2. Cheng, R. and J. Vassileva, *Design and evaluation of an adaptive incentive mechanism for sustained educational online communities.* UMUAI, 2006. **V16**(3): p. 348.
3. Falkowski, T. and M. Spiliopoulou, *Users in Volatile Communities: Studying Active Participation and Community Evolution*, in *User Modeling*. 2007. p. 47-56.
4. Fischer, G. and J. Ostwald, *Knowledge Management: Problems, Promises, Realities, and Challenges.* IEEE Intelligent Systems, 2001. **16**(1): p. 60 - 72.
5. Freeman, L., *Centrality in social networks conceptual clarification.* Social Networks, 1979. **1**(3): p. 239.
6. Harper, M., et al. *Talk amongst yourselves: inviting users to participate in online conversations*. in *IUI 2007*: ACM.
7. 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.
8. Kay, J., et al., *The Big Five and Visualisations of Team Work Activity*, in *Intelligent Tutoring Systems*. 2006. p. 197-206.
9. Author Citation Deleted for Anonymity Reasons.
10. Author Citation Deleted for Anonymity Reasons.
11. Lave, J. and E. Wenger, *Situated Learning: Legitimate Peripheral Participation*. 1991, New York: Cambridge University Press.
12. Seco, N., T. Veale, and J. Hayes. *An Intrinsic Information Content Metric for Semantic Similarity in WordNet*.
13. Song, X., et al., *ExpertiseNet: Relational and Evolutionary Expert Modeling*. Lecture Notes in Computer Science. 2005. 108.
14. 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.
15. 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.